Fusion and Inference of Geometric Information and Functional Contrast in Computational Anatomy

by

Jianqiao Feng

A dissertation submitted to The Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy.

Baltimore, Maryland

October, 2014

Abstract

Magnetic resonance imaging (MRI) has been widely used as a noninvasive clinical and research modality for the study of human anatomy. In the past decade, Computational Anatomy (CA) has emerged as a discipline to study the neuroanatomical variability via morphometric mapping algorithms. Quantitative analysis of anatomy has thus become possible.

This dissertation discusses feature extraction, fusion and inference of geometric information and functional contrast from MRI scans in the computational anatomy framework. An important application of the methodology presented here is the diagnosis of human brain neurodegenerative diseases, e.g. Alzheimer's diseases. Through this dissertation we consider the problem of distinguishing between healthy controls (HC) and Alzheimer's disease (AD).

The Human brain with its highly complex anatomy is composed of a number of subregions, or subcortical structures. It is known that different diseases affect different regions of human brain. As a result, this dissertation focuses on regions of interests (ROI), i.e. subcortical structures instead of analysis of whole brain. To capture

ABSTRACT

morphological changes of a certain subcortical structure affected by AD, a surface-based statistical analysis is firstly presented. This approach is extended to multiple structure analysis to combine discriminative information from different structures. The result shows that different structures carry complementary information.

Besides the geometric feature, functional contrast feature can be extracted and added to the classification procedure. All 3D structural images are transformed into a common template coordinate system. Jacobian of deformation field and intensity value at each voxel are used as geometric and functional contrast features respectively. The feature selection is performed to avoid potential over-training. Data fusion methods are employed to combine feature vectors of different categories extracted from different structures. This data analysis pipeline is validated using a public medical image database. Higher performance is observed compared with that using geometric feature alone, which indicates fusion of geometric and functional contrast can improve classification performance. A comparable or higher classification accuracy is achieved compared with two state-of-the-art methods. Moreover, different from previous methods, our methods provide a way to extract biomarkers easily interpretable.

Primary Reader: Dr. Michael I. Miller

Herschel and Ruth Seder Professor, Department of Biomedical Engineering, Johns Hopkins University

ABSTRACT

Director, Center for Imaging Science, Johns Hopkins University

Professor, Department of Electrical and Computer Engineering, Johns Hopkins University

Secondary Reader: Dr. John Goutsias

Professor, Department of Electrical and Computer Engineering, Johns Hopkins University

Acknowledgments

I still remember that morning in the summer four years ago. It is the first time I came to Dr. Miller's office and discuss with him whether we could find some opportunities to work together. Five minutes later we made up the decision that I would work at the Center for Imaging Science (CIS) for a few months. Three months later, he agreed that we should work together during my PhD program and be my advisor. The most impressive for me is his optimism, every time I was frustrated he always encourages me and give lots of valuable suggestions to me.

I would also thank Dr. Priebe for encouraging and helping me. I learned a lot in his statistical classes. Many years later, this would become a very good memory that every Wednesday morning I struggled to get up very early to catch up the meeting with him.

I am grateful to all those professors on the committee board of my dissertation defense and graduate oral exam, Dr. Goutsias, Dr. Tran, Dr. Dredze, Dr. Naiman, and Dr. Younes.

I would like to thank my labmates, Xiaoying Tang, Daniel Tward, Valentina

ACKNOWLEDGMENTS

Staneva, Saurabh Jain, and Kwame Kutten for their discussions on diffeomorphism, geometry, and manifold learning. I want to express special thanks the staffs at CIS, Mike Bowers, Anthony Kolasny, Dawn Kilheffer, Maura Vonasek, Timothy Brown, and Erika Lance for making CIS a sweet home. Many thanks to all my friends who have accompanied me through the years in Baltimore, especially those from Peking University.

Last but not least, I would like to thank my parents, Lixin Feng and Wenxiang Chen, for their love and support through my PhD program. When I decided to come to U.S. for pursuing a doctoral degree, they supported my decision without any hesitation, although none of them has ever been this country before. And many thanks to my girlfriend, Zheng Ma. Thank you for believing in me, and making me a better man.

The work reported here is supported by grants: R01 EB000975, P41 EB015909, and R01 EB008171.

Abbreviations

MRI: Magnetic Resonance Imaging

CA: Computational Anatomy

AD: Alzheimer's Disease

ADNI: Alzheimer's Disease Neuroimaging Initiative

HC: Healthy Control

ROI: Regions of Interests

SVM: Support Vector Machine

MKL: Multiple Kernel Learning

MK-SVM: Multi-Kernel Support Vector Machine

LDDMM: Large Deformation Diffeomorphic Metric Mapping

MC-LDDMM: Multi-Channel Large Deformation Diffeomorphic Metric Mapping

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 Overview

Today, brain MRI (magnetic resonance imaging) scans have become a widely used clinical test for human brain disease diagnosis. MRI based study of neuroanatomical variability of the human brain has been employed in various neurodegenerative diseases, normal aging, and neuropsychiatric disorders [1].

In the field of Computational Anatomy (CA) [2], morphometric mapping algorithms are developed to study the neuroanatomical variability. This dissertation is devoted to extraction, fusion, and inference of geometric information and functional contrast of structural images via computational anatomy methodologies. Instead of studying the whole brain, here we only consider seven regions of interest (ROI), or subcortical structures of the brain. To be specific, they are amygdala, caudate, hip-

pocampus, pallidus, putamen, thalamus, and lateral ventricle. The motivation is the fact that different structures are responsible for a range of functions, and diseases can generally be associated with geometric and functional contrast abnormalities in structural images. For example, Alzheimer's disease (AD) leads changes in the temporal lobe.

In particular, through the dissertation we focus on Alzheimer's disease, i.e. to discriminate healthy control (HC) subjects and AD patients by analysis of certain subcortical structures extracted from the baseline MRI scan, i.e. the first scan.

The dissertation is organized as follows:

Chapter 1 presents an overview and introduce the dataset used through this dissertation. Besides some background knowledge about human brain anatomy, computational anatomy model and machine learning technologies is provided. Besides, some related works are introduced in the end of this chapter.

Chapter 2 demonstrates a surface based method for identifying AD from normal aging. A metric distance is measured using surface mapping algorithm. Thus a metric space is defined for anatomical structures. Manifold learning technology is employed to reduce the dimensionality.

Chapter 3 extends the method introduced in Chapter 2 from single structure to multiple anatomical structures. In order to utilize geometric information extracted from multiple structures simultaneously, kernel based methods are applied for inference and information fusion.

Chapter 4 discusses fusion of geometric information and functional contrast. These two kinds of information is extracted via 3D image registration.

Chapter 5 concludes the dissertation, describing some ongoing work and potential future directions.

All proposed methods are evaluated on a dataset obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) [3]. More details about the dataset we used in experiments are presented in section 1.3.

1.2 Subcortical Structures

In this dissertation we study on seven subcortical structures of the human brain on both left and right sides. They are amygdala, caudate, hippocampus, pallidum, putamen, thalamus, and ventricle. Figure 1.1 visualizes the iso-surfaces of these structures with different structures shown in different colors.

The **amygdala** is an ovoid shaped subcortical structure located in the dedial temporal lobe, inferior to the putamen and anterior to the hippocampus. It is responsible for the formation and storage of memories of emotional events. Figure 1.2 and 1.3 visualize examples of amygdalae on both left and right sides to present the location of the amygdala in human brain and the shape.

The **caudate** is an subcortical structure located in the dedial temporal lobe, inferior to the lateral ventricles and superior to the putamen. The caudate plays an

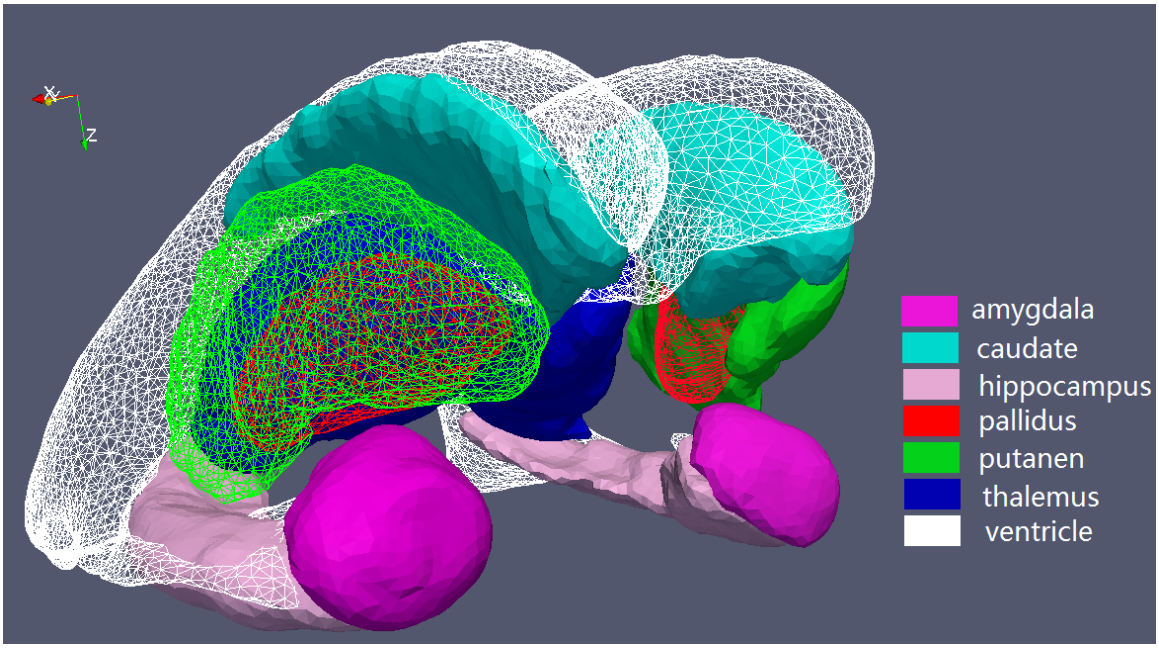


Figure 1.1: All subcortical structures in surface representation.

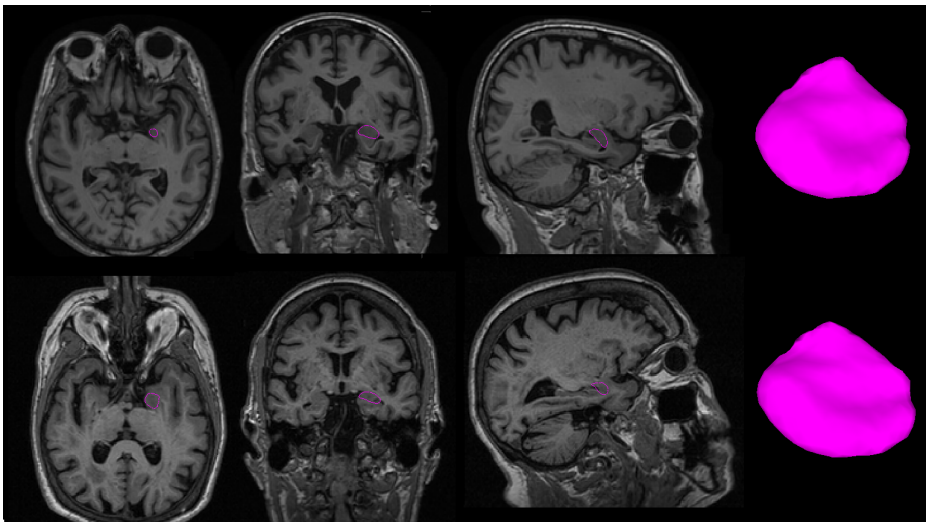


Figure 1.2: Two examples of the left amygdala. For each row, from left to right: axial slice, coronal slice, sagittal slice with purple contour at the boundary of left amygdala, and surface representation of left amygdala. Two examples are shown in two rows respectively.

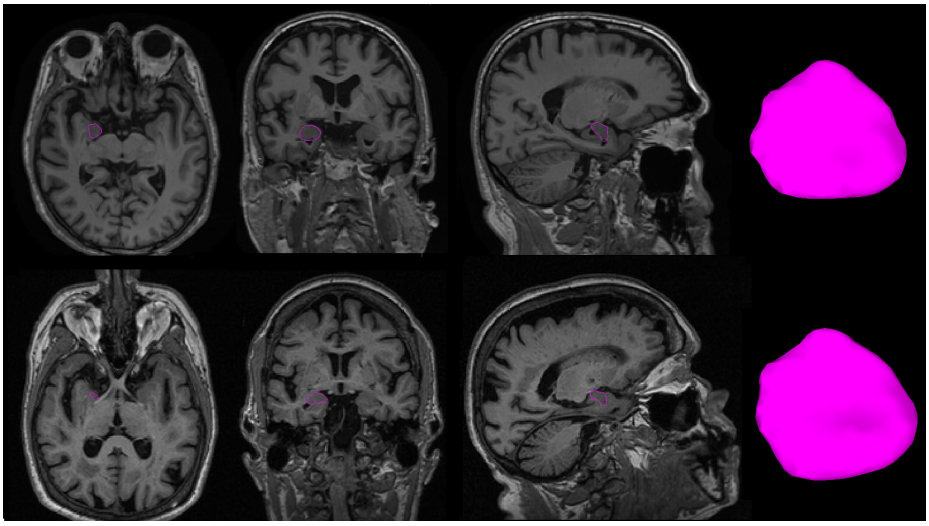


Figure 1.3: Two examples of the right amygdala. For each row, from left to right: axial slice, coronal slice, sagittal slice with purple contour at the boundary of right amygdala, and surface representation of right amygdala. Two examples are shown in two rows respectively.

important role in voluntary movement, learning, memory, sleep, and social behavior. Figure 1.4 and 1.5 visualize examples of the caudate on both left and right sides to present the location and the shape.

The **hippocampus** is an horseshoe shaped subcortical structure located in the temporal lobe, posterior to the amygdala. It plays a central role in long-term memory. Figure 1.6 and 1.7 present examples of hippocampi on both left and right sides to show the location of the hippocampus in human brain and the shape.

The **pallidum** is an ovoid shaped subcortical structure located between the thalemus and the putamen. It is associated with the regulation of voluntary movement. Figure 1.8 and 1.9 visualize examples of pallidum on both left and right sides to demonstrate the location in human brain and the shape.

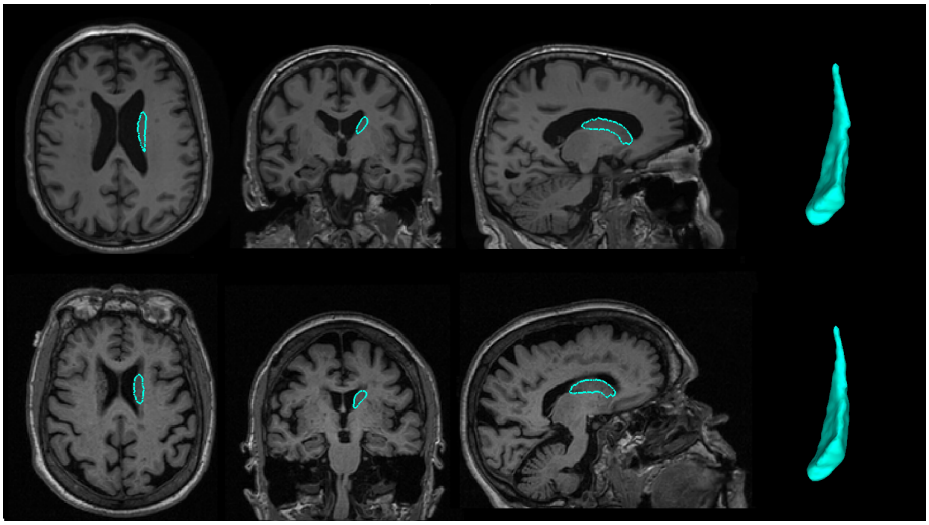


Figure 1.4: Two examples of the left caudate. For each row, from left to right: axial slice, coronal slice, sagittal slice with light blue contour at the boundary of left caudate, and surface representation of left caudate. Two examples are shown in two rows respectively.

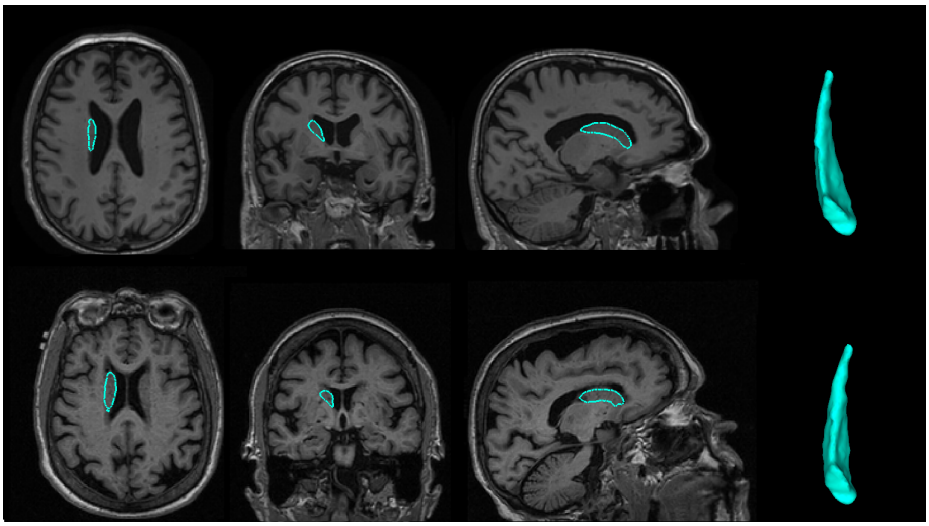


Figure 1.5: Two examples of the right caudate. For each row, from left to right: axial slice, coronal slice, sagittal slice with light blue contour at the boundary of right caudate, and surface representation of right caudate. Two examples are shown in two rows respectively.

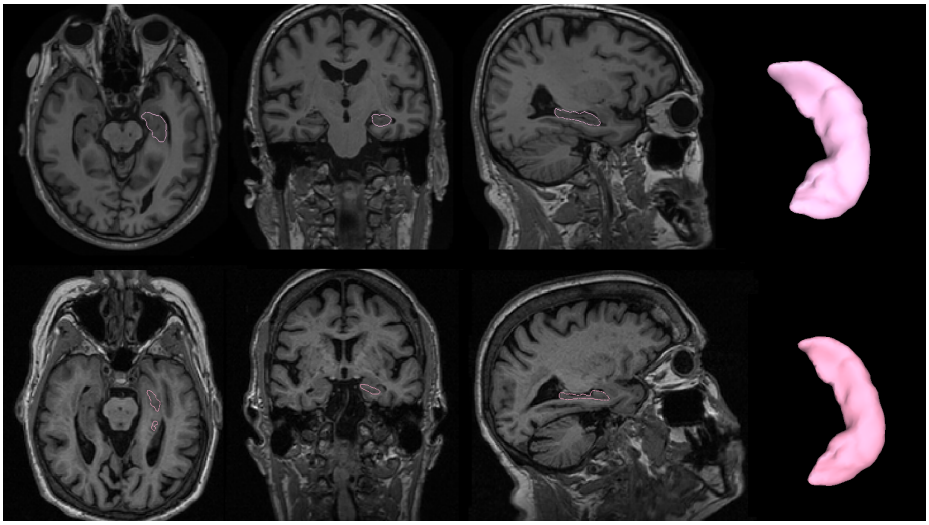


Figure 1.6: Two examples of the left hippocampus. For each row, from left to right: axial slice, coronal slice, sagittal slice with pink contour at the boundary of left hippocampus, and surface representation of left hippocampus. Two examples are shown in two rows respectively.

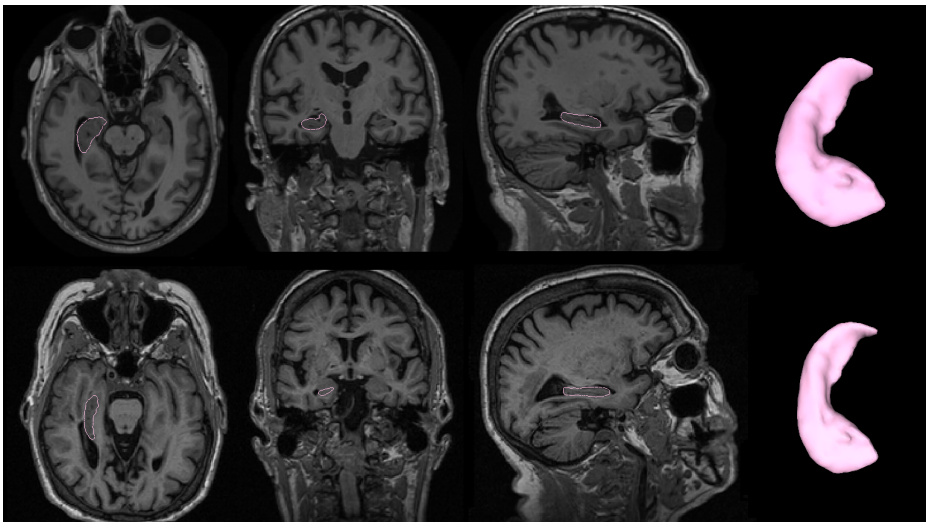


Figure 1.7: Two examples of the right hippocampus. For each row, from left to right: axial slice, coronal slice, sagittal slice with pink contour at the boundary of right hippocampus, and surface representation of right hippocampus. Two examples are shown in two rows respectively.

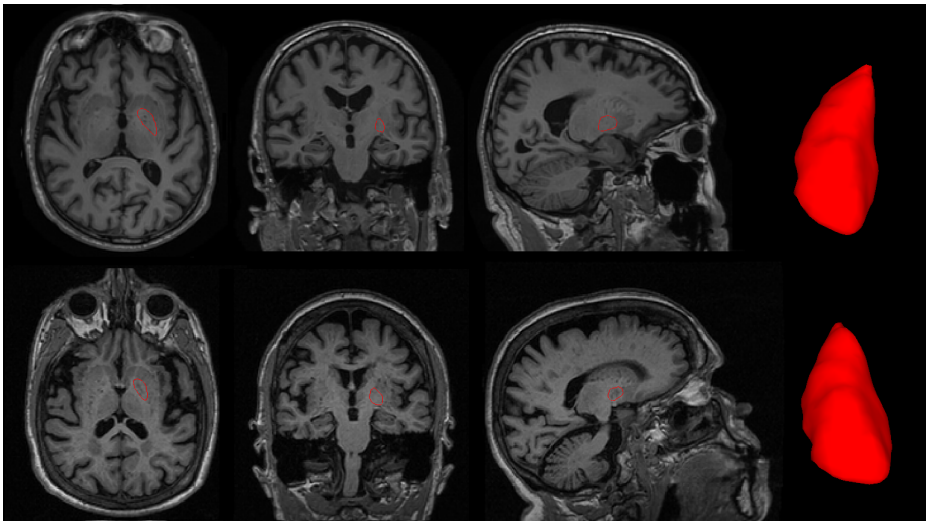


Figure 1.8: Two examples of the left pallidum. For each row, from left to right: axial slice, coronal slice, sagittal slice with red contour at the boundary of left pallidum, and surface representation of left pallidum. Two examples are shown in two rows respectively.

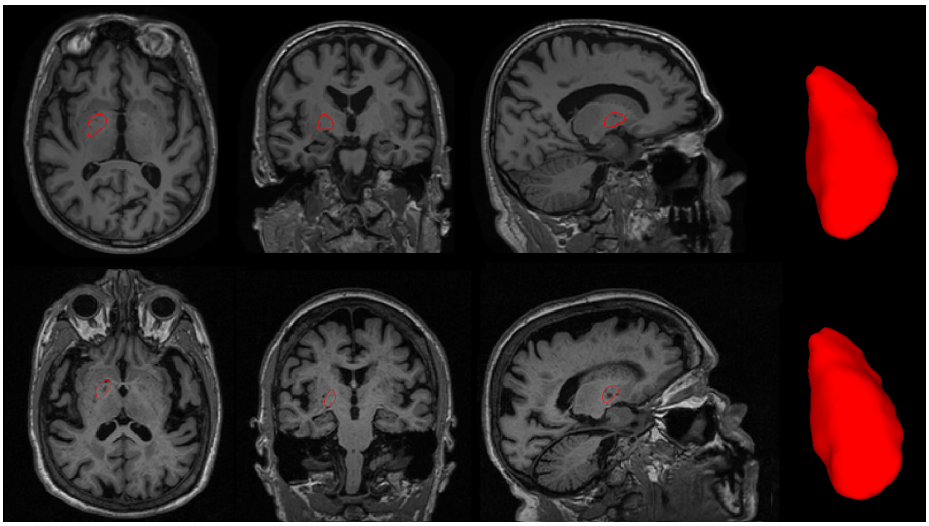


Figure 1.9: Two examples of the right pallidum. For each row, from left to right: axial slice, coronal slice, sagittal slice with red contour at the boundary of right pallidum, and surface representation of right pallidum. Two examples are shown in two rows respectively.

The **putamen** is a round subcortical structure which plays a role in motor skills and many types of learning. Figure 1.10 and 1.11 visualize examples of putamen on both left and right sides to show the location in the human brain and its shape.

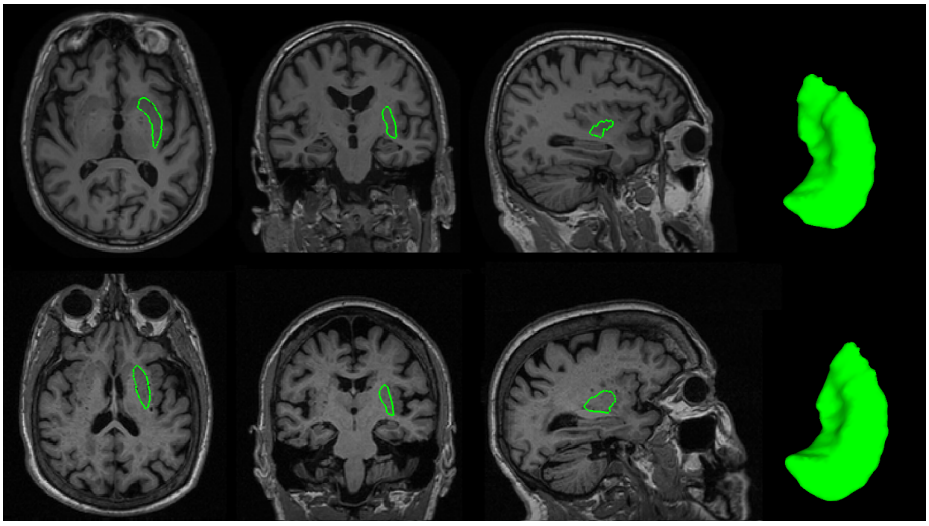


Figure 1.10: Two examples of the left putamen. For each row, from left to right: axial slice, coronal slice, sagittal slice with green contour at the boundary of left putamen, and surface representation of left putamen. Two examples are shown in two rows respectively.

The **thalamus** is a subcortical structure located within the vertebrate brain, surrounded by the lateral ventricle. It is involved in the regulation of consciousness, sleep, and motor skills. Figure 1.12 and 1.13 present examples of the thalamus on both left and right sides to demonstrate the location of the thalamus in human brain and the shape.

The **lateral ventricles**, the largest ventricles in human brain, are C-shaped subcortical structures containing cerebrospinal fluid (CSF) located within vertebrate brain. The volume of ventricles is of clinical significance for diagnosis of a variety

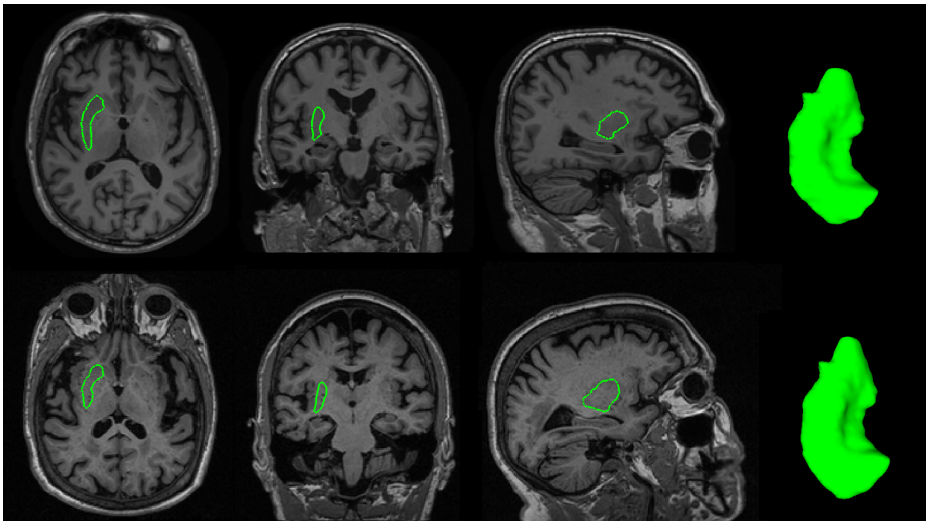


Figure 1.11: Two examples of the right putamen. For each row, from left to right: axial slice, coronal slice, sagittal slice with green contour at the boundary of right putamen, and surface representation of right putamen. Two examples are shown in two rows respectively.

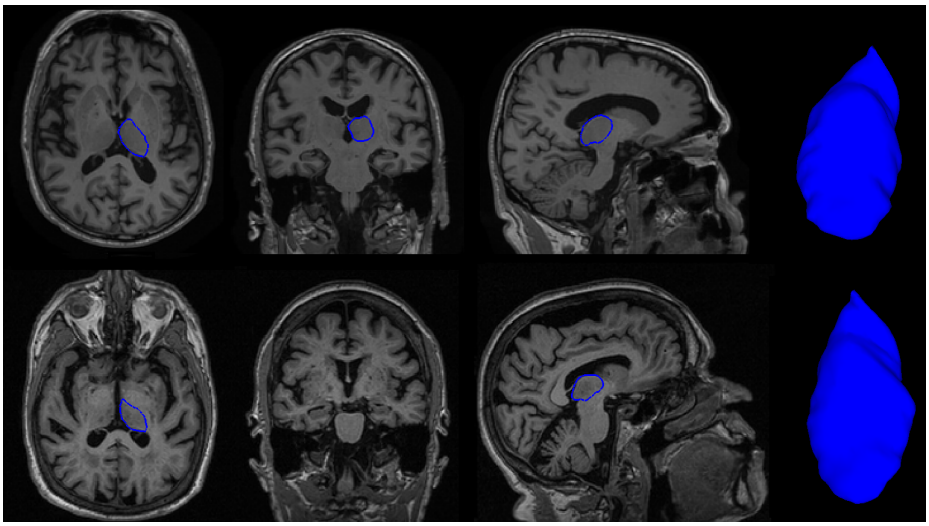


Figure 1.12: Two examples of the left thalamus. For each row, from left to right: axial slice, coronal slice, sagittal slice with blue contour at the boundary of left thalamus, and surface representation of left thalamus. Two examples are shown in two rows respectively.

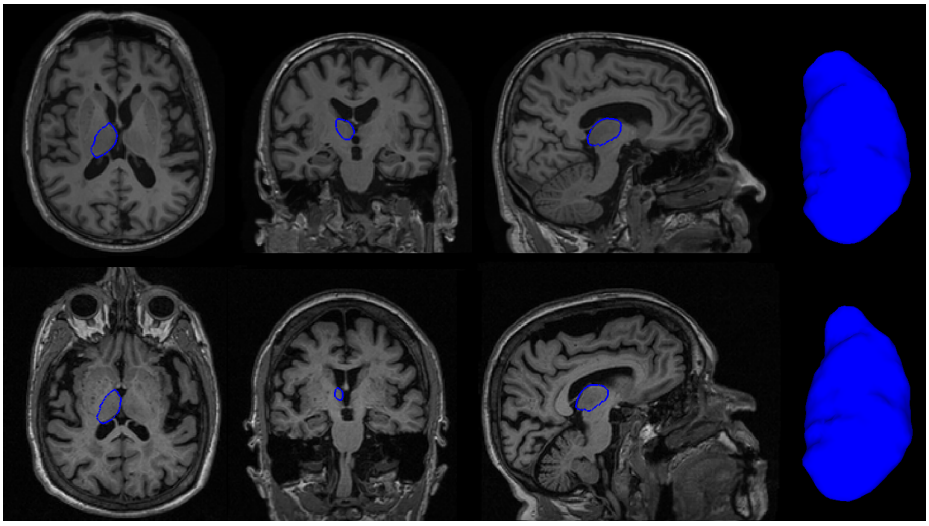


Figure 1.13: Two examples of the right thalamus. For each row, from left to right: axial slice, coronal slice, sagittal slice with blue contour at the boundary of right thalamus, and surface representation of right thalamus. Two examples are shown in two rows respectively.

of neurological disorders, e.g. Alzheimer's disease and schizophrenia. Figure 1.14 and 1.15 present examples of the lateral ventricles on both left and right sides to show the location of lateral ventricles in human brain and the shape.

1.3 Data

The data used in all experiments reported in this dissertation are obtained from ADNI (adni.loni.ucla.edu), which was launched in 2003 by the Food and Drug Administraion (FDA), the National Institute of Biomedical Imaging and Bioengineering (NIBIB), the national Institute on Aging (NIA), private companies, and non-profit organizations.

Diagnosis of AD especially at early stage is very chanllenging. The primary goal

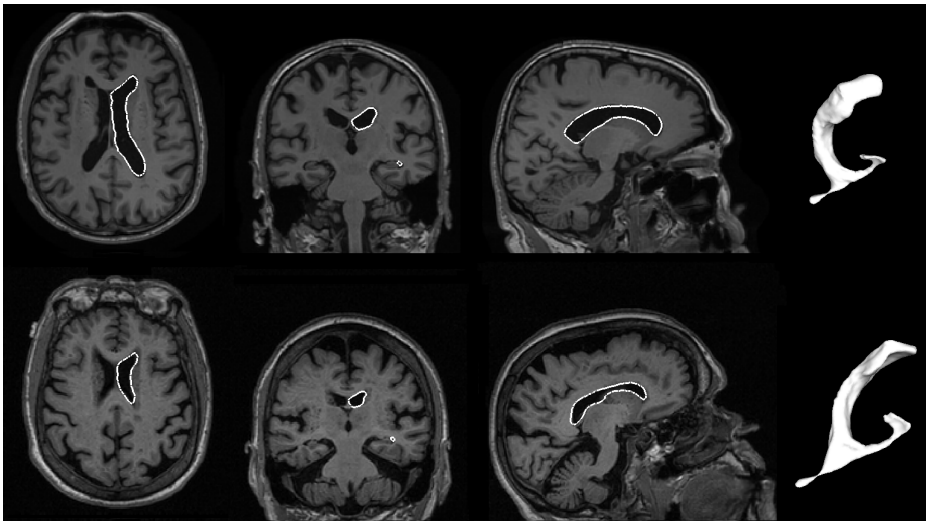


Figure 1.14: Two examples of the left lateral ventricle. For each row, from left to right: axial slice, coronal slice, sagittal slice with white contour at the boundary of left lateral ventricle, and surface representation of left lateral ventricle. Two examples are shown in two rows respectively.

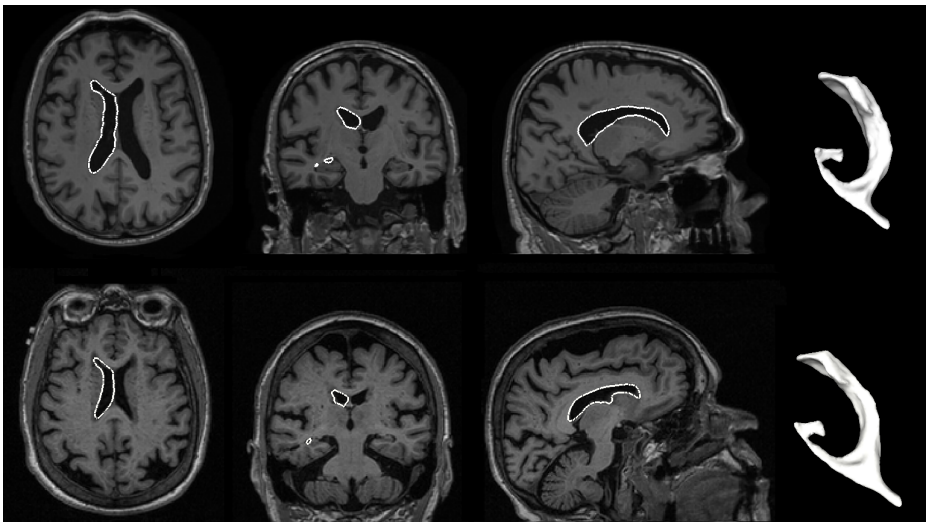


Figure 1.15: Two examples of the right lateral ventricle. For each row, from left to right: axial slice, coronal slice, sagittal slice with white contour at the boundary of right lateral ventricle, and surface representation of right lateral ventricle. Two examples are shown in two rows respectively.

of ADNI is to test whether the progression of mild cognitive impairment (MCI) can be measured by combining different imaging modalities, such as MRI and Positron Emission Tomography (PET), other biological markers, and neuropsychological and clinical assessment. Ultimately ADNI is aimed at improving effectiveness and efficiency of clinical trials via determination of sensitive and specific biomarkers for progression of AD. Over 800 subjects ages between 55 and 90 have been recruited from more than 50 sites across the U.S. and Cannada in the first 5 years since its launch. The principal investigator is Michael W. Weiner from University of California, San Francisco and VA Medical Center. For up-to-date information, visit the website www.adni-info.org.

The dataset used in our experiment contains 385 T1-weighted MR images obtained from ADNI. Although there are more than 800 subjects with 4000 scans in ADNI database, we only considered the healthy control and disease groups, with the baseline, i.e. the first scan, for each subject. Some scans (84 out of 840) were excluded because they suffered severe degradation due to motion artifacts or significant clinical abnormalities (e.g., hemispheric infarction). A dataset of 756 subjects was formed after this unbiased selection, including 210 subjects of HC, 175 subjects of AD, and 371 subjects of Mild Cognitive Impairment (MCI). Table 1.1 presents detailed information of this dataset.

Table 1.1: Demographic characteristics of the dataset used in this paper

Group	HC	MCI	AD
number of subjects	210	369	175
number of male subjects	109	236	94
Age (year)	76.25 ± 5.01	75.03 ± 7.32	75.28 ± 7.49

1.3.1 Segmentation of Subcortical Structures and Surface Generation

For each subject, there are two raw T1-weighted images. In order to improve signal-to-noise ratio, these two images are rigid-body aligned to a common coordinate and averaged. Another preprocessing is to resample images to isotropic 1mm voxels. The T1 image of the i-th subject after preprocessing is denoted as T_i, where $i = 1, \ldots, N$. Here $N = 385$ is the number of subjects in database.

The T1-weighted image T_i is segmented into several subcortical structures via a 3D image segmentation software FreeSurfer [4, 5]. There are seven subcortical structures, hippocampus, amygdala, caudate, putamen, globus pallidus, thalamus, and lateral ventricle are studied (both left and right sides) in this dissertation. Figure 1.16 shows an example of T1 weighted scan and segmentation created via FreeSurfer.

To guarantee the quality of segmentations, a quality review of segmentations was carried out. Three technicians were recruited, who had at least four months of experience in reviewing brain MR images. After training from an expert neuroanatoimst, qualitative review was performed by one of these technicians.

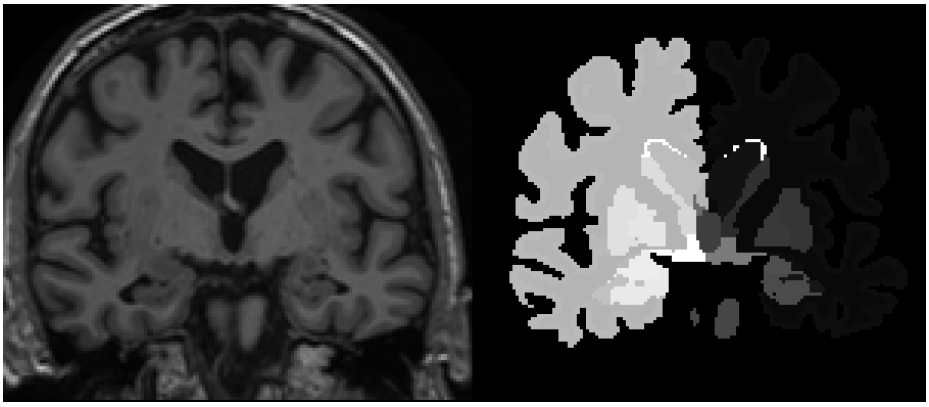


Figure 1.16: A coronal slice of a T1 weighted MRI scan (left) and segmentation (right) generated via Freesurfer.

As described in [6, 7, 8] images that suffered degradation were excluded. Possible cause of such degradation includes artifacts caused by motion and technical problems, and significant clinical abnormalities.

Surface based shape analysis can better capture subtle change of anatomical shape compared to volumetric analysis. Thus Smooth 2D Surfaces contouring the corresponding volume are generated from volumetric segmentation for shape analysis. The quality of triangulated surfaces generated from volumetric segmentation using marching cubes [9] is not satisfying. As a result a pipeline built on LDDMM is adopted. The Computational Functional Anatomy (CFA) subcortical template is a template set which is created from 41 manually labeled volumes. This template set is composed of seven structures as mentioned above on both sides [10, 1]. The properties of template surface, like smoothness and correct topology, is carried to the deformed template surface through diffeomorphic mapping [11]. For the template image, there are 14 binary images (7 structure on both sides) and the corresponding contouring

surfaces are denoted as $I_{temp} = \{I_{temp}^{(s)}\}$ and $S_{temp} = \{S_{temp}^{(s)}, \}$, where $s = 1, \ldots, M$ is the index of structures. In our case $M = 14$ which is the number of structures. Each I_{temp}^{s} is a binary image with 1 indicate the position is inside the subcortical structure. For the i-th subject, denote the segmentation images $J_i = \{J_i^{(s)}\}$, where $s = 1, \ldots, M$ and $i = 1, \ldots, N$. In our case $N = 385$ is the number of subjects in the dataset. The segmentation images were created using FreeSurfer [5, 4]. The template segmentation images are transformed towards segmentation image of each subject with transformation $\phi_1^{(i)}$ calculated via LDDMM algorithm. The deformed template image $\hat{J}_i$ is an approximation of the original segmentation image with less noise. Without introducing ambiguity, the same notation J_i is used for the segmentation after this denoising preprocessing.

$$\hat{J}_i = \{\hat{J}_i^{(s)}\} = \{I_{temp}^{(s)} \circ (\phi_1^{(i)})^{-1}\} \tag{1.1}$$

The smooth contouring surfaces S_i^s are generated from these segmentation images. The surface based method demonstrated in Chapter 1 is validated using these S_i^s. This method of surface generation has already been validated with examples to show discrepancies between $\hat{I}_{sub_i}$ and original I_{sub_i}. Another quantitative validation [12] shows the volume difference is within 10% for most of structures.

1.4 The Basic Model of Computational Anatom

Computational Anatomy (CA) is a discipline to study of biological variability of human anatomy [2]. In CA framework, the anatomical variability is understood by studying the diffeomorphism ϕ mapping anatomical mainfolds to one another. A diffeomorphism is a smooth invertible transformation which is defined everywhere on the background space. The manifolds could be landmarks (1D), surfaces (2D), and structural images (3D). Without loss of generality, let's take images as an example. Given a template image I_{temp}, formally the anatomy is modelled as a triple $(\Omega, \mathcal{G}, \mathcal{I})$, where Ω is the background space, $\mathcal{G}$ is a group of diffeomorphisms defined on Ω, $\mathcal{I}$ is the orbit of the template image I_{temp} under $\mathcal{G}$. So $\mathcal{I}$ is the set of all images generated by applying a certain diffeomorhpism on the template image. In equation 1.2, $I_{temp} \circ \phi^{-1}$ means applying the diffeomorphism ϕ on the template image.

$$\mathcal{I} = \{I : I_{temp} \circ \phi^{-1}, \forall \phi \in \mathcal{G}\} \tag{1.2}$$

The flow of diffeomorphism ϕ_t mapping template image I_{temp} to target image J, is connected with a underlying velocity field, a flow of smooth time-dependent vector field $v_t \in V, t \in [0, 1]$, by a ordinary differential equation 1.3. The end point of the

flow, ϕ_1 is the diffeomorphism mapping the template image to the target image.

$$\dot{\phi}_t = v_t(\phi_t), \phi_0 = Id, t \in [0,1] \ , \tag{1.3}$$

For any pair of target images $I, J \in \mathcal{I}$, we can define a distance between I and J as below. It has been proved this is a metric distance between I and J [13].

$$\begin{aligned} \rho^2(I,J) \ &\triangleq \inf_{\dot{\phi}_t = v_t(\phi_t), \phi_0 = Id} \int_0^1 \|v_t^*\|_v^2 dt, \\ &\quad such\ that\ \phi_0 \cdot I_{temp} = I, \phi_1 \cdot I_{temp} = J. \end{aligned} \tag{1.4}$$

Here $\|v_t\|_v$ is a norm defined by a operator L (equation 1.5). The operator $L = Id + \alpha\nabla^2$, where Id is identity, and ∇^2 is the Laplacian operator which enforce some smoothness of the velocity field and thus guarantee the ϕ_1 is a diffeomorphism [14]. The varying parameter α controls the smoothness of the velocity field and deformation. A smoother deformation is generated with a larger α. And with a smaller α, a more accurate transformation is calculated.

$$\|v_t\|_v = \|Lv_t\|_2^2 = < L^\dagger L v_t, v_t >_{l^2} \tag{1.5}$$

In general we cannot guarantee that the condition $J \in \mathcal{I}$ always holds. Instead we use the random orbit model, which modelled the target image as a sum of image on the orbit and a noise signal, i.e., $J = I^D = I + noise$, where $I \in \mathcal{I}$. Thus this

becomes an inexactly matching problem. The $l-2$ norm can be used to measure the mismatching error, i.e. $\|J - I_{temp} \circ \phi_1^{-1}\|^2$. The large deformation diffeomorphic metric mapping (LDDMM) is thus proposed to solve this inexactly matching problem. As a widely used tool of computational anatomy, it seeks the optimal velocity field v_t^* to minimize a energy function, defined by Eq. 1.6, combining smoothness and mismatching error of the mapping ϕ with a trade off parameter σ.

$$\begin{aligned} v_t^* &= \arg\min_{v \in V} E(v) \\ &= \arg \min_{v \in V, \dot{\phi}_t = v_t(\phi_t), \phi_0 = Id, \phi_1 = \int_0^1 v_t(\phi_t) dt} \int_0^1 \|v_t\|_v^2 dt + \frac{1}{\sigma^2} D(I \circ \phi_1^{-1}, J) \end{aligned} \quad (1.6)$$

An example of 2D image registration is shown in figure 1.17. Figure 1.18 shows transformed template image over the flow of diffeomorphism ϕ_t, i.e., $I_t = I_{temp} \circ \phi_t^{-1}$, t=$\{$0,0.2,0.3,0.5,0.7,0.8,1$\}$.

1.5 Background of Machine Learning Technologies

In this dissertation, a number of machine learning technologies are adopted. This section briefly introduces some basic concepts in machine learning field. Machine learning is a set of algorithms try to extract knowledge or to learn rules from dataset $D = \{(x_i, y_i) | x_i \in R^m, y_i \in R, i = 1, \ldots, n\}$. The x_i is the data sample, or the

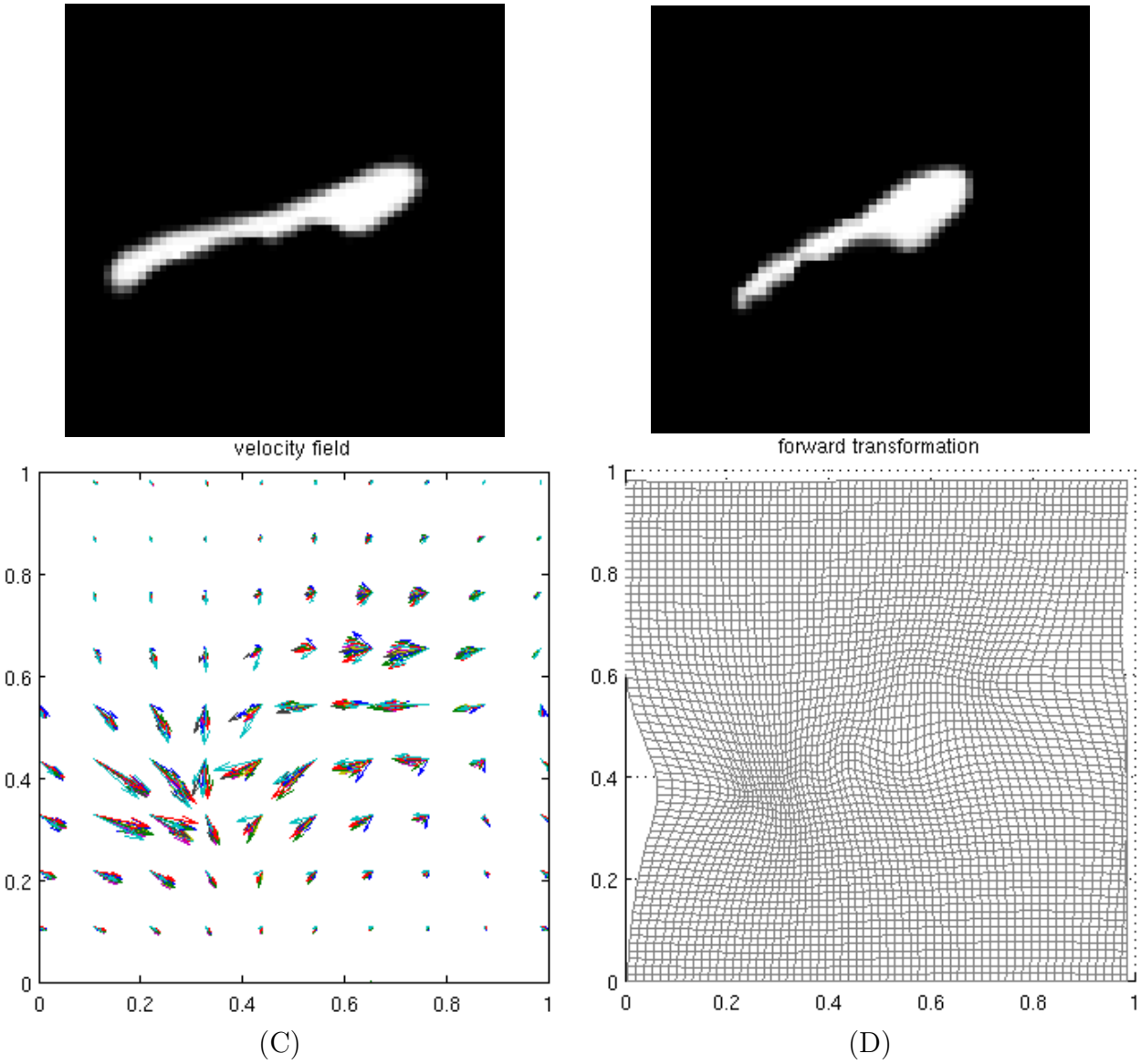


Figure 1.17: An example of input and output of LDDMM algorithm. (A) the template image I_{temp}. (B) the target image J. (C) the velocity field v_t, different color indicate the velocity field at different time point. (D) the deformation field ϕ_1.

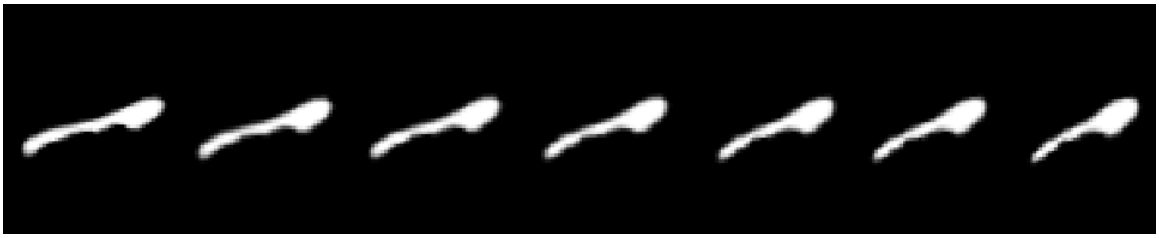


Figure 1.18: Transformed template over the flow of diffeomorphism, i.e. $I_t = I_{temp} \circ \phi_t^{-1}$, t=$\{0,0.2,0.3,0.5,0.7,0.8,1\}$.

feature vector of a data sample with **dimensionality** equal to m. The y_i is the corresponding **label**. The n is the size of dataset. The task of machine learning algorithms is to find a function of x, $\hat{y} = f_{\hat{\theta}}(x)$, to predict y, where θ is the parameter of the model. The data samples used for estimating parameters form the **training set**. The **testing set** is composed of data samples x used for testing the performance of $f_{\hat{\theta}}(x)$.

1.5.1 Categories

Supervised learning algorithms train the model on labeled data examples, i.e., y_i is available for all x_i in training set. The supervised learning algorithm attempts to generalize a trained prediction function from training set to test set, i.e. to generate an predicted label for previously unseen data samples speculatively. Depending on whether the label is continuous or discrete, supervised learning algorithms can be categorized into **classification** (discrete label) and **regression** (continuous label). This dissertation focus on the earlier one.

Unsupervised learning algorithms work on unlabeled examples, i.e., y_i is un-

available for each x_i in training set. Unlike supervised learning, the goal of unsupervised learning algorithms are not to seek a mapping from feature vector to the label. Instead the objective is to discover structure in the data, e.g. clusters in the data in clustering algorithms.

Semi-supervised learning algorithms are trained on dataset partially labeled. There are other categories left out of the discussion of this dissertation, e.g. Reinforcement learning, Transduction, etc.

1.5.2 Evaluation of Classification Performance

Because only classification algorithms are considered throughout this dissertation, here several criteria to evaluate the performance of classification algorithms are introduced. In context of medicine screening or disease diagnosis, the class label is $\{Positive, Negative\}$. True Positive (TP) is the number of positive subjects predicted as positive correctly. By similar way, we can define False Positive (FP), False Negative (FN), and True Negative (TN). The relationship between these concepts are demonstrated in table 1.2. True Positive Rate (TPR), True Negative Rate (TNR), **precision** or Positive predictive value (PPV), Negative predictive value (NPV), and **accuracy** are defined as below. Accuracy, or equivalently **Misclassification Rate (MCR)** measures the possibility of correct prediction. **Sensitivity** (true positive rate) and **specificity** (true negative rate) measure the proportion of positive and

negative samples being correctly predicted respectively.

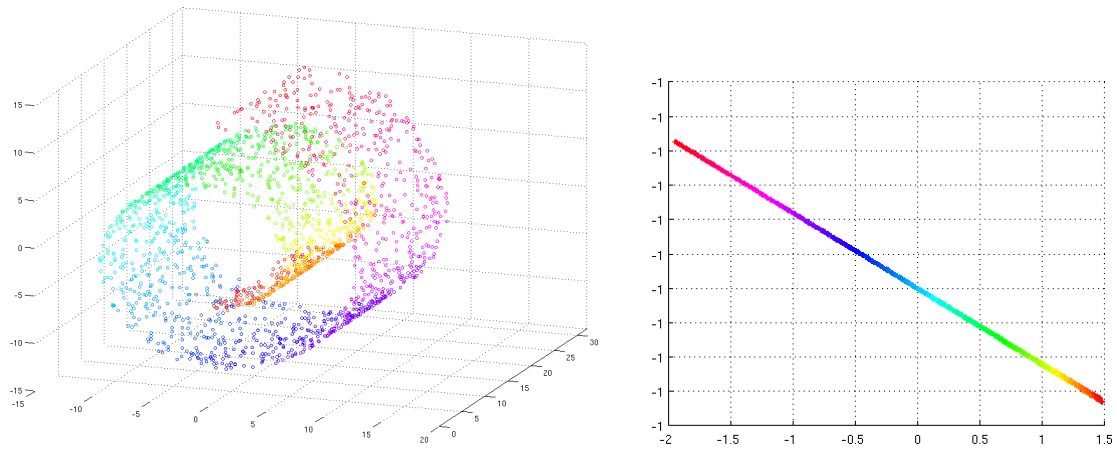
$$
\begin{aligned}
Accuracy(ACC) &= 1 - MCR = \frac{TP + TN}{TP + FP + FN + TN} \\
Sensitivity &= TPR = \frac{TP}{TP + FN} \\
Specificity(SPC) &= TNR = \frac{TN}{FP + TN} = 1 - FPR \\
PPV &= precision = \frac{TP}{TP + FP} = \frac{TP}{P} = TPR \\
NPV &= \frac{TN}{TN + FN}
\end{aligned}
$$

Table 1.2: Confusion Matrix

		Prediction		
		Positive	Negative	Rate
Ground Truth	Positive	True Positive (TP)	False Negative (FN)	TPR
	Negative	False Positive (FP)	True Negative (TN)	TNR
	Rate	PVV	NPV	

1.5.3 Manifold Learning

High-dimensionality is a big challenge in machine learning field. Thus dimension reduction algorithms have been developed to seek a low-dimensional representation for high-dimensional feature vectors. Manifold learning is one of such approaches based on the assumption that data points lie on an embedded non-linear manifold within the higher-dimensional space. Unlike dimension reduction methods that provide mappings from original feature space to low dimensional space, manifold learning methods



(A) (B)

Figure 1.19: A demonstration of manifold learning algorithm. (a) data samples visualized in original 3D features space. (b) data samples embedded in a 2D space, with embeddings calculated via Isomap.

typically calculate low-dimensional representation of data samples to approximate a distance measurements. Formally, given a N by N distance matrix, where N is the number of objects or data samples, manifold learning algorithms aim to place each data sample in N-dimensional space such that the pairwise distances are preserved as well as possible. Each object is then assigned coordinates, i.e. the embeddings, in the N-dimensional space.

Classical Multidimensional Scaling (MDS) takes an input matrix, $D_{i,j}$ and seeks a coordinate matrix, $x_i \in R^N, i = 1, \ldots, N$. The problem is to approximate $D_{i,j}^2$ by $\|x_i - x_j\|^2$, which can be transformed to approximation of inner product. The inner product data $B_{i,j}$ is derived from distance data $D_{i,j}$: $B_{i,j} = \tilde{D_{i,j}} - \tilde{D_{\cdot,j}} - \tilde{D_{i,\cdot}} + \tilde{D_{\cdot,\cdot}}$, where $\tilde{D}_{i,j} \triangleq -D_{i,j}^2/2$ and $\tilde{D_{\cdot,j}}$, $\tilde{D_{i,\cdot}}$, $\tilde{D_{\cdot,\cdot}}$ are row, column and grand means of $\tilde{D}$. The

loss function of classical MDS to be minimized is called strain [15] is defined by 1.7.

$$Strain_D(x_1, \ldots, x_N) = \left(\frac{\sum_{i,j} (B_{i,j} - < x_i, x_j >)^2}{\sum_{i,j} B_{i,j}^2} \right)^{1/2} \tag{1.7}$$

Let E_m be the matrix of m eigenvectors and Λ_m be the diagonal matrix of m eigenvalues of B. The coordinate matrix X is calculated by $X = E_m \Lambda_m^{\frac{1}{2}}$.

Isomap extends classical MDS. It assumes that the linearity only holds in a local neighborhood. The distances are set as unknown except for neighboring points. This incomplete matrix is restored using the FloydWarshall algorithm [16] to compute the shortest-path distances between all pairs. The embeddings are calcuated from the restored distance matrix using MDS. The solution is obtained from the eigen-value decomposition of B. An example is demonstrated in figure 1.19. In the original 3D space, data samples lie on a 2D manifold, which is known as "swiss roll". The data points are plot out using the first two dimensions of embeddings corresponding to the two largest as coordinates.

Laplacian Eigenmap (LLE) is graph based manifold learning algorithm [17]. The embeddings are calculated by eigen-value decomposition of graph Laplacian L defined by 1.9. The first step is to construct a graph with each vertex represents a data sample. The edges are only put between close vertexes, which therefore are connected. A weight matrix is calculated, e.g. using heat kernel $W_{i,j} = exp\{-\frac{D_{i,j}^2}{\sigma}\}$

where σ is a parameter.

$$Lf = \quad \lambda Df \tag{1.8}$$

$$L = \quad D - W \tag{1.9}$$

Let $f_0, \ldots, f_k$ be the solution of the generalized eigen-vector problem defined in 1.8, corresponding to the eigen-values in descending order. The embeddings of for each data point is $x_i = (f_1(i), \ldots, f_m)$. Note the first eigen-vector f_0 is skipped.

1.6 Previous Work

In literature, many effective methods were proposed based on different biomarkers, i.e., MRI, functional imaging (e.g., PET), with some successful application of combination of these different modalities [18]. However in this dissertation, we focus on one single modality, i.e., MRI. A brilliant survey to compare classification methods of automatically discriminating between patients with AD or mild cognitive impairment (MCI) and healthy controls based on T1-weighted MR images from ADNI database was published [19].

These methods can be roughly divided into three categories according to the type of features extracted from MR images for classification. From large to small scale: 1) volumetric measurements of regions of interest (ROI) like hippocampi. Methods of this category are based on hippocampal volume [20, 21] or more accurate shape anal-

ysis of hippocampi [22], 2) vertex-based methods. These methods extract features at the vertex-level on cortical surfaces, e.g., cortical thickness [23, 24]. and 3) voxel-wise features [25, 26]. This kind of method defines features at each voxel of 3D MR images. Classification methods capable of dealing with high dimensional data like SVM are employed. Another option is to reduce the dimensionality of the feature space using dimension reduction techniques like PCA and/or feature selection methods.

In this dissertation, three methods are discussed. The methods are combination of first and third categories as discussed above. The analysis was carried out on ROIs, i.e., subcortical structures. But features were extracted at voxel-level from structural images. Specifically we focused on seven subcortical structures (on both left and right sides), amygdala, caudate, hippocampus, pallidum, putamen, thalamus, and ventricle. In the first two surface based methods, morphological variation is extracted from different structures. The third one extracts voxel-wise geometric and photometric features from seven structures.

Chapter 2

Surface Based Metric Space Structure Analysis

2.1 Introduction

This chapter describes a method based on metric structures for anatomical analysis on a large set of brain MR images. In the past decade, Computational Anatomy (CA) [2] has emerged as a discipline to quantitatively analyze the neuroanatomical variability via morphometric mapping algorithms. In this chapter, a widely-used framework in CA, Large Deformation Diffeomorphic Metric Mapping (LDDMM) [27] was used for dissimilarity measurement. A geodesic distance between each pair of subcortical structures was measured using LDDMM. Manifold learning approaches were applied to seek a low-dimensional representation of healthy control (HC) and

Alzheimer's disease (AD) subjects in the high-dimensional shape space. Inference between healthy and disease groups can be done using traditional classification algorithms. The proposed method was evaluated on ADNI, a dataset for the Alzheimer's disease study as claimed in section 1.3. Instead of studying the whole brain, here we consider four anatomical structures, amygdalae, caudate, hippocampi, and ventricles which have been reported to be affected morphologically by AD. Our work demonstrates that the high-dimensional anatomical shape space of these four structures can be approximated by a relatively low dimension manifold.

There have been several related methods proposed in the literature to apply manifold learning methods to dissimilarities measured on structural images. Most of them quantify inter-subject dissimilarity based on volume image data. For example, a similarity measurement called bending energy is employed in [28]. In [29, 30], random forests have been used to measure dissimilarity on some statistical region-based features extracted from volume images. In [31], dissimilarity was calculated via a small deformation to approximate a large deformation for computational efficiency. Similar approximations can be found in [32, 33, 28]. However, the explosive growth of computing power along with parallel computing resources have made the problem of computation less severe. Thus a full large deformation diffeomorphic metric is adopted in this paper. In this study, we couple the volume imagery to surfaces, and track the variations of shape using surface models. Studying surfaces allows us to capture the variation of neurodevelopment very efficiently [32].

Low-dimensional embeddings can be calculated from the dissimilarity information via standard manifold learning approaches, such as multidimensional scaling (MDS) used in [29, 30, 34], Isomap used in [28, 32, 33], and Laplacian Eigenmaps (LE) used in [31].

After the embedding in low-dimensional space is calculated, statistical inference is often carried out. One choice is classification between different cohorts. Classification accuracy then becomes an important criteria for evaluating the dimension of the embedding. We note that we only consider the two-class classification problem, i.e., HC and AD. In this paper, a number of standard manifold learning and classification algorithms were employed. Section 2.3 presents our experiments and results. Compared with related works mentioned above, our investigations are more extensive, i.e., we consider a larger dataset. Our methodology achieves comparable classification accuracy.

2.2 Method

In this section, the framework is described step by step. The first part of the framework is to calculate a distance matrix whose element is pairwise dissimilarity, as presented in Figure 2.1. For manifold learning, data samples are embeded into a low dimensional space and classified in that space, shown in Figure 2.2.

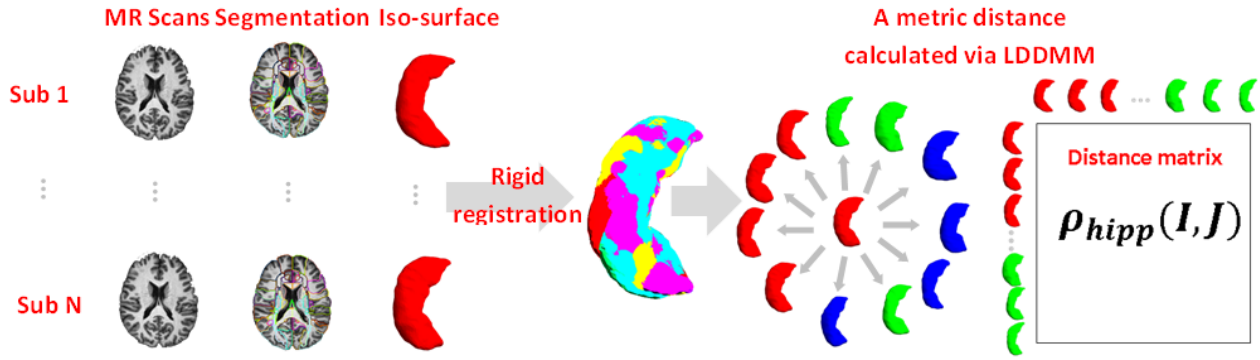


Figure 2.1: The flowchart of the framework: data acquisition, segmentation, surface extraction, rigid transformation, and LDDMM-surface mapping.

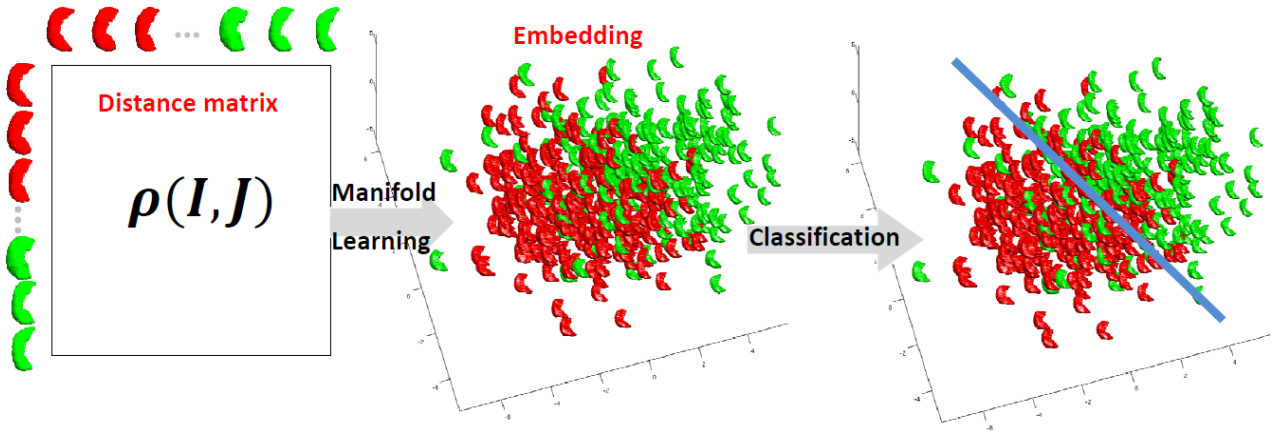


Figure 2.2: The flowchart of the framework: embedding calculation and classification (one structure).

2.2.1 Preprocessing: Segmentation and Iso-surface extraction

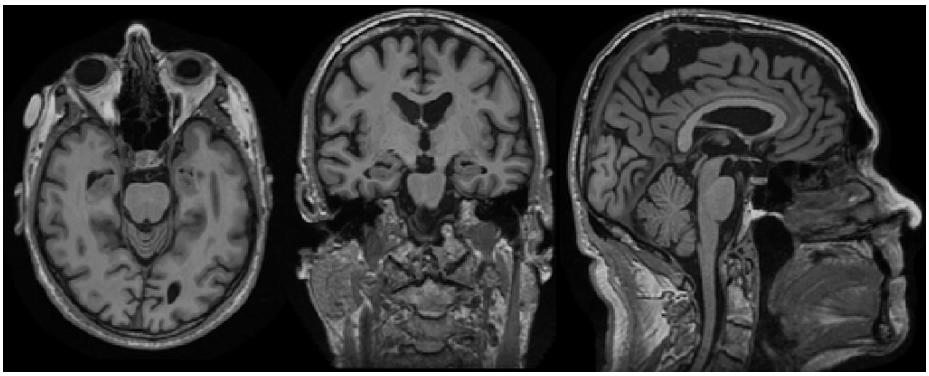


Figure 2.3: An example of healthy control. From left to right: axial slice, coronal slice, sagittal slice of the T1 weighted MRI scan.

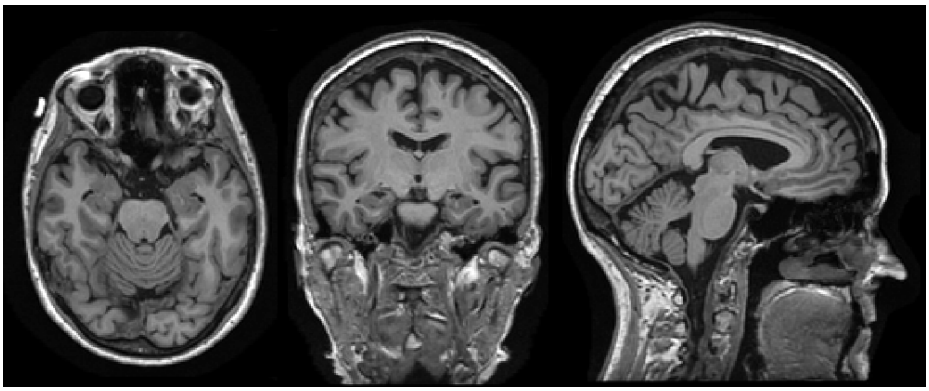


Figure 2.4: An example of Alzheimer's Disease patient. From left to right: axial slice, coronal slice, sagittal slice of the T1 weighted MRI scan.

Figure 2.3 and 2.4 present T1 images of a healthy control and Alzheimer's Disease patient respectively. As detailed in section 1.3.1, we followed a similar procedure as used in [1], in which template surfaces are used to initialize the topology of the subcortical structures upon which the inference will be performed, and then tar-

get surfaces are inserted into the segmentations from Freesurfer [4, 5] via LDDMM mapping. These template surfaces transformed towards different target surfaces are accurate since the LDDMM mappings used are smooth, and mediate the noise which may be sometimes inherent in segmentations. These surfaces then become the manifolds that our inference is based on. Figure 2.5 presents some examples of iso-surfaces of hippocampi from AD and HC group.

2.2.2 Rigid Alignment

Prior to LDDMM, a rigid alignment [12] was carried out to remove the variation caused by different poses in image acquisition stage. Given two surface $S_{template}$ and S_{target} with sets of vertex $x_i, i = 1, \ldots, M$ and $y_i, i = 1, \ldots, N$ respectively. The method is similar as the one proposed in [35]. It seeks an optimal transformation parameters, rotation matrix $R(\theta)$ and translation vector T to minimize the cost function defined by Eq. 2.1, where the first term $3 - tr(R(\theta))$ is the regularization term. The second and third terms are the matching and soft assign terms. Figure 2.6 presents

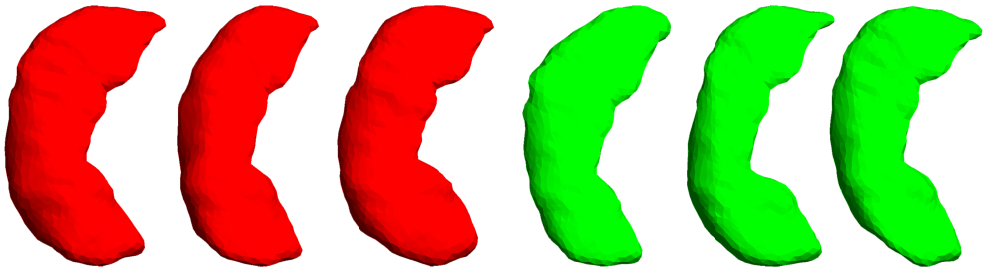


Figure 2.5: Some examples of iso-surfaces of hippocampi with AD in red and HC in green.

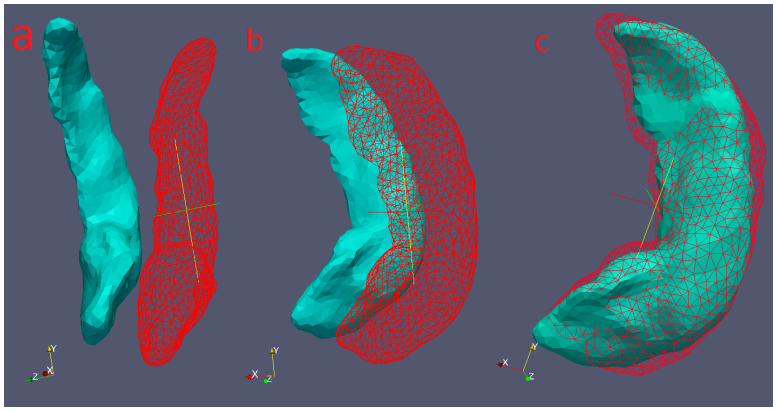


Figure 2.6: Example of surface rigid registration result. Panel a and b are template (blue) and target surface (red) before rigid registration. Panel c visualizes the transformed template surface and target surface.

an example of rigid registration result.

$$\begin{aligned} E(\theta, T) &= \lambda(3 - tr(R(\theta))) + \sum_{i=1}^{M}\sum_{j=1}^{N}\{(W_{i,j} + v_{ij})\|R(\theta)x_i + T - y_j\|^2 \\ &\quad + t(w_{ij}logw_{ij} + v_{ij}logv_{ij})\} \\ &\quad \sum_{i=1}^{M} = 1, \sum_{i=1}^{M} = 1, w_{ij} \geq 0, v_{ij} \geq 0 \end{aligned} \tag{2.1}$$

2.2.3 Diffeomorphic Metric Mapping

To measure dissimilarity between two subjects, diffeomorphic metric mapping is calculated via LDDMM surface mapping [36, 37]. The diffeomorphism ϕ_t is constructed as a flow using the ordinary differential equation:

$$\dot{\phi}_t = v_t(\phi_t), \phi_0 = Id, t \in [0, 1] \ , \tag{2.2}$$

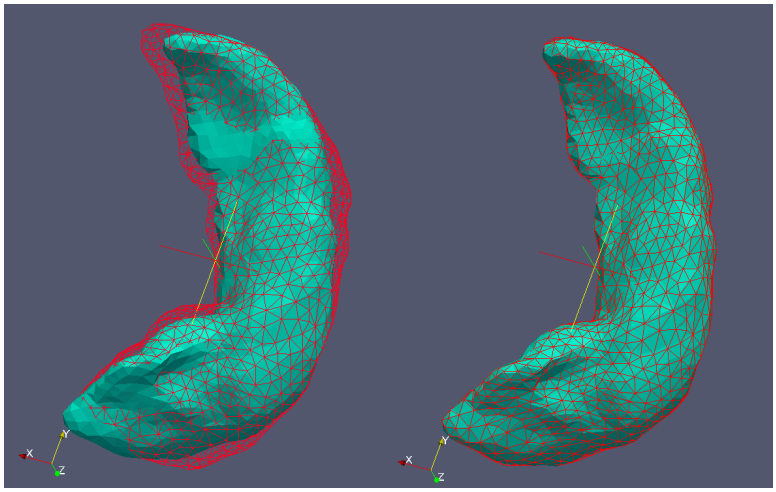


Figure 2.7: Example of LDDMM surface registration result. Left picture is the template (blue) and target surface (red) after rigid registration. Right one shows the result of transformed template surface and target surface via LDDMM surface registration.

where v_t is the velocity vector field which determines the corresponding flow ϕ_t. The boundary value of this ODE is the identity map denoted as Id. Figure 2.7 presents an example of LDDMM surface registration result.

Given a pair of surfaces I and J, a dissimilarity $\rho^2(I, J)$ between them is calculated by integrating the norm of velocity vector field associated with the geodesic ϕ_t over time, where σ is the parameter for trade-off between smoothness and goodness of fit. The LDDMM surface mapping algorithm seeks the optimal time dependent velocity field v_t^* to minimize a loss function, defined by Eq. 2.3, combining smoothness and goodness of fit of the mapping ϕ.

$$\begin{aligned} v_t^* &= \arg\min_{v \in V} E(v) \\ &= \arg\min_{v \in V} \int_0^1 \|v_t\|_v^2 dt + \frac{1}{\sigma^2} D(\phi_1 \cdot I, J) \end{aligned} \tag{2.3}$$

The first integral term measures the smoothness of velocity field v_t. The smoothness ensures v_t lies in a smooth Hilbert space $(V, \|\cdot\|_v)$ and the generated ϕ_t is a diffeomorphic transformation. The second term is the mismatching term measuring discrepancy between the transformed template and the target surface. For surface mapping this discrepancy term $D(S_{template}, T_{target})$ can be expressed in a mixture of inner products between norms of faces from $S_{template}$ and S_{target}. Let f , g are the faces of $S_{template}$ and q, r are faces of S_{target}. As defined in [36] the mismatching term can be written as Eq. 2.4 show, where $c(\cdot)$ and $N(\cdot)$ are the center and normal vector of a face on a triangulated surface, and k_w is a kernel function in the form of a 3X3 diagonal matrix, e.g. $k_w = diag(e^{-\|x-y\|^2/(2\sigma^2)}, e^{-\|x-y\|^2/(2\sigma^2)}, e^{-\|x-y\|^2/(2\sigma^2)})$.

$$\begin{aligned} D(S_{template}, T_{target}) = \quad & \sum_{f,g} N(f)^t k_w(c(g), c(f)) N(g) \\ & -2 \sum_{f,q} N(f)^t k_w(c(q), c(f)) N(q) \\ & + \sum_{q,r} N(q)^t k_w(c(q), c(r)) N(r) \end{aligned} \tag{2.4}$$

The metric distance between I and J is given by Eq. 2.5.

$$\rho^2(I, J) = \int_0^1 \|v_t^*\|_v^2 dt \tag{2.5}$$

Note that the second term in the cost function $E(v)$, goodness of fit, implies this is not an exact matching problem, because the exact matching problem does not have

a well defined solution. Thus $\rho(I, J)$ is not a precise metric for it is not symmetric, and this effect has been studied in [38]. Here we took the "averaging" strategy , i.e. $\hat{\rho}(I, J) = \hat{\rho}(J, I) \triangleq \frac{(\rho(I,J)+\rho(J,I))}{2}$, to make it symmetric.

- "averaging" strategy: $\hat{\rho}(I, J) = \hat{\rho}(J, I) \triangleq \frac{(\rho(I,J)+\rho(J,I))}{2}$
- "Min" strategy: $\hat{\rho}(I, J) = \hat{\rho}(J, I) \triangleq min\{\rho(I, J), \rho(J, I)\}$
- "max" strategy: $\hat{\rho}(I, J) = \hat{\rho}(J, I) \triangleq max\{\rho(I, J), \rho(J, I)\}$

2.2.4 Manifold Learning and Classification

Manifold learning is a popular approach of non-linear dimension-reduction. It attempts to find a low-dimensional embedding (i.e., the manifold) in the high-dimensional space. The hypothesis is that the data points are samples from a low-dimensional manifold. There are a number of algorithms in the manifold learning family. Here we consider three of them, i.e., classical MDS, Isomap [39], and Laplacian Eigenmaps (LE) [40].

Four classification algorithms, K-Nearest Neighbor (KNN) and Fisher's linear discriminant analysis (LDA), Support Vector Machine (SVM), and Random Forest (RF) were employed to discriminate HC and AD cohorts.

The Random forest method[41] is an extension of traditional decision tree, which makes an overall prediction based on decisions of all individual trees.

Here we give a brief introduction to SVM, because in chapter 3 the kernel SVM is extended to multiple kernel SVM for multiple structure analysis. Here we take the two-class C-Support Vector Classification (C-SVC) [42] as an example. Given training samples $x_i \in R^n, i = 1, \ldots, l$, in two classes. The ground truth of class label $y \in R^l$, where $y_i \in \{1, -1\}$. In principle, the goal of the SVM method is to seek an optimal hyperplane to maximize the margin. An example of 2D is given in figure 2.8.

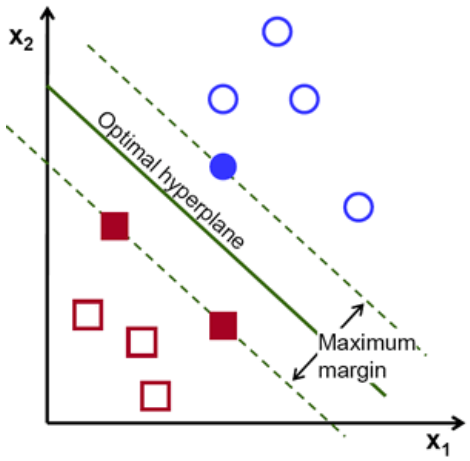


Figure 2.8: Demonstration of SVM.

Formally, the primal optimization problem of C-SVC is defined as follows:

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^{l} \xi_i \\ & y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, \ldots, l. \end{aligned} \tag{2.6}$$

where the vector of weight coefficients $w \in R^d$ and the bias term b are the pa-

rameters to specify the hyperplane, ξ is the slack variable to handle the case that data samples of two classes are not linearly separable. The feature mapping function $\phi : R^n \to R^d$ transform a data sample from original feature space of n dimension to new feature space of d dimension (d could be very large, even infinite). The predefined parameter C is a trade-off parameter between training error and model simplicity.

Usually, we solve the dual problem, especially in case of extremely high dimensional feature space,

$$\begin{aligned} \min_{\alpha} \quad & \mathcal{S}(\alpha) = \frac{1}{2}\alpha^T Q\alpha - e^T\alpha \\ & 0 \le \alpha_i \le C, \qquad i = 1, \ldots, l, \\ & y^T\alpha = 0, \end{aligned} \tag{2.7}$$

where $e = [1, \ldots, 1]^T$ is the vector of all ones, C is the upper bound of α, Q is an l by l positive semidefinite matrix, $Q_{ij} \equiv y_i y_j K(x_i, x_j)$, and $K(x_i, x_j) \equiv \phi(x_i)^T\phi(x_j)$ is the kernel. Thus kernel trick is utilized for the case of infinite dimensional feature space.

The optimal w could be calculated from α via $w = \sum_{i=1}^{l} y_i\alpha_i\phi(x_i)$. The decision function is defined as $\hat{y} = f(x) = sign(\sum_{i=1}^{l} y_i\alpha_i K(x_i, x) + b)$.

Thanks to the kernel trick, one could directly define a kernel matrix without explicitly providing feature mapping function ϕ. Here we consider four different kernels, linear kernel (SVM_L), polynomial kernel (SVM_P, with degree p), radial

basis function (SVM_R), and sigmoid (SVM_S). The formal definitions are provided below.

$$K_L(x_i, x_j) = \quad x_i^T x_j =< x_i, x_j > \tag{2.8}$$

$$K_P(x_i, x_j) = \quad (< x_i, x_j > +b)^p \tag{2.9}$$

$$K_R(x_i, x_j) = \quad exp(-\gamma \|x_i - x_j\|^2) \tag{2.10}$$

$$K_S(x_i, x_j) = \quad tanh(\alpha < x_i, x_j > +\beta) \tag{2.11}$$

An open source implementation of SVM, LIBSVM was used [43, 44] for all experiments presented in this dissertation.

2.3 Experiments and Result

We tested the proposed method on four anatomical structures, the amygdala, caudate, hippocampus, and ventricle. For each subject, eight surfaces including both left and right sides are studied. On any of these surfaces, an inter-subject dissimilarity was calculated via LDDMM following a rigid registration as described in section 2.2.2 and 2.2.3.

Figure 2.9 demonstrates an example of the dissimilarity matrix, which is calculated from the left hippocampus. Each pixel represents an entry of the distance

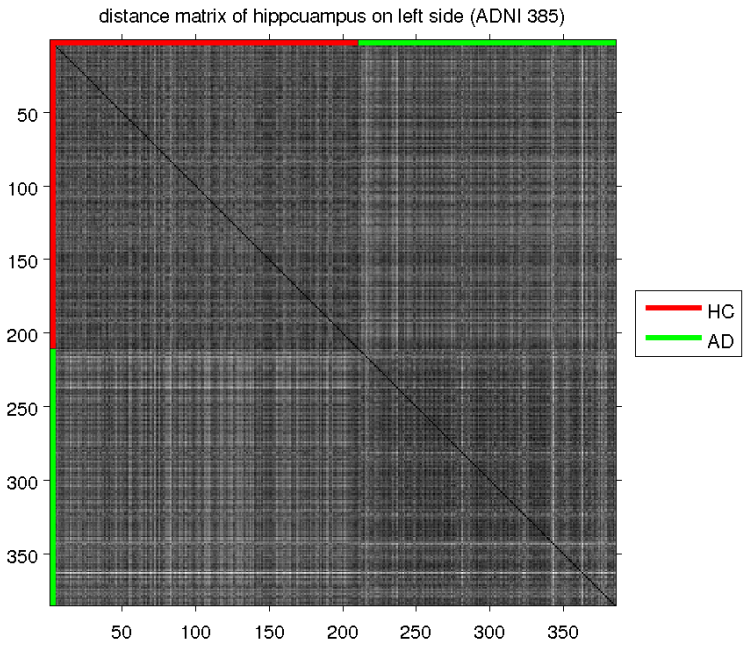


Figure 2.9: Distance matrices calculated on left hippocampus, using LDDMM-surface mappinge).

matrix $D^s_{i,j}, s = 1, \ldots, S$, at (i, j), which is the metric distance between hippocampi of subject i and j calculated via LDDMM. The superscript s on distance matrix $D^s, s = 1, \ldots, S$ is the index of anatomical structures. For example we have four structures, hippocampus, amygdala, caudate, and ventricle on both left and right sides. In total we have eight distance matrices. So the number of structure $S = 8$ here. Matrices are of size 385 by 385, with 210 healthy controls, and the other 175 AD patients.

For convenience of visualization, all entries are normalized into the range of $[0, 1]$. The brighter it looks, the larger distance it represents. From figure 2.9, one can observe that two diagonal blocks are darker than two off-diagonal blocks. In other

words, the metric distance measured within the class is smaller than that across different classes in principle. This make sense because shape difference should be larger between subjects from different classes. Four examples of hippcampi and amygdalae

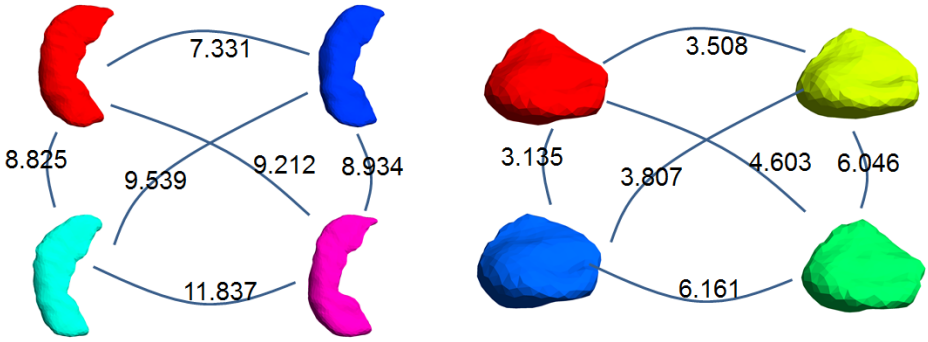


Figure 2.10: Distance matrices Calculated on left hippocampus and left amygdala using LDDMM-surface registration. The number on the line connecting two structures is the metric distance between them.

are randomly selected from the dataset. The metric distance measured between each pair is shown in Figure 2.10.

Given the set of surface maps, one of three manifold learning methods, MDS, Isomap, and Laplacian Eigenmaps, was employed to calculate embedding, on which several classifiers were trained and tested. The embeddings calculated from left hippocampus using MDS is visualized in figure 2.11. Each surface of hippocampus is placed at (x, y, z), which are the first three dimensions corresponding to three largest eigen-values. Healthy control samples are shown in the red color, while the green ones are from AD group.

In general the AD and HC groups are separable on the low-dimensional space calculated via manifold learning methods.

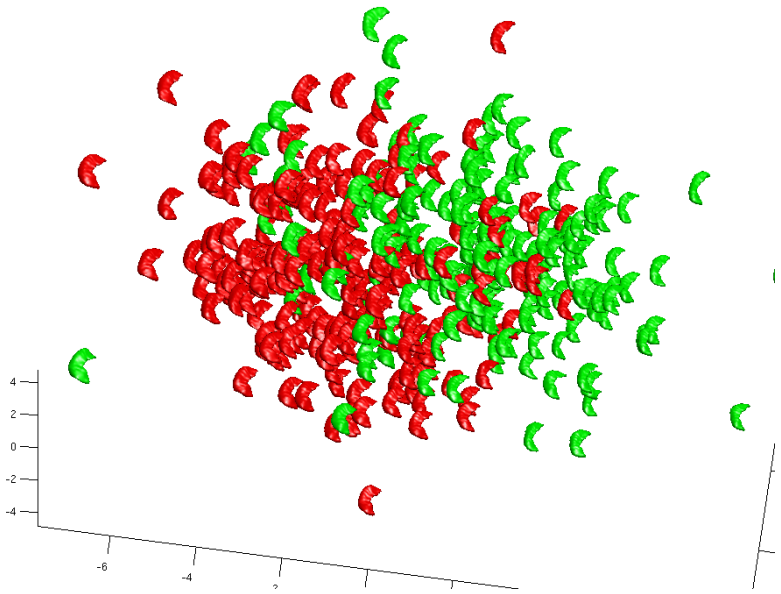
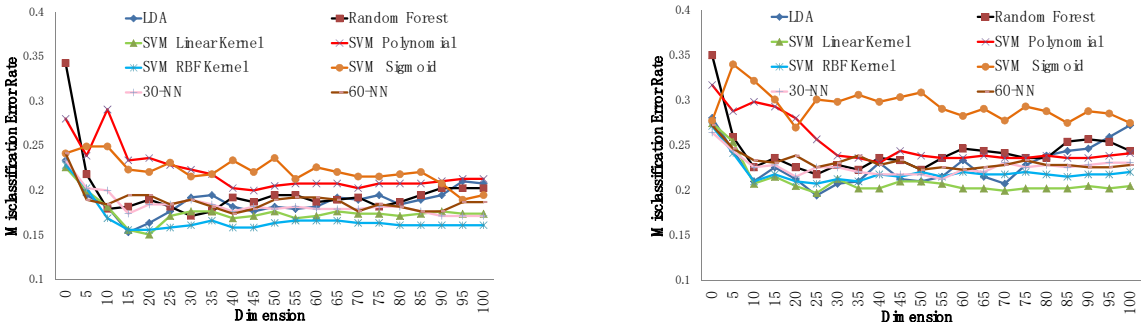


Figure 2.11: The first 3 dimensions of embeddings calculated via MDS (left hippocampus).



(A) (B)

Figure 2.12: Misclassification rate as a measure of the embedding dimensions from the hippocampus via MDS. (A) is the result of left hippocampus. (B) is the result of right hippocampus.

A 5-fold cross-validation was carried out to evaluate the performance of classification on 385 subjects (210 HC and 175 AD). The performance was measured in terms of misclassification rate.

In order to understand whether the dimension of representation for the anatomical shape space is large or small, misclassification rate was calculated by varying the dimensions of the embedding. In particular, we considered the first d dimensions corresponding to the largest d eigen-values for $d = 1, ...90$. As a result, a curve of misclassification rates over different dimensions was obtained for each classification algorithm. Figure 2.12 presents misclassification rates, using MDS followed by certain classification algorithms, as a function of the number of embedding dimensions calculated from the left hippocampus. From the misclassification curves, one can observe that the misclassification rate first decreases rapidly then increases slowly, as the embedding dimensions increases. Thus, a low dimensional embedding is sufficient to achieve a low misclassification rate. However, increasing the dimensionality can lead to noises and over-fitting [45, 46]. The dimensions with lowest misclassification rate under our setup, is 15 to 20 for almost all of the classification algorithms approximately. This suggests that anatomical shape lies in a space of relatively low dimensionality.

Similar results can be observed for the other two manifold learning methods, Isomap and Laplacian Eigenmaps, shown in figure 2.13 and 2.14.

Table 2.1 is a full comparison of minimum misclassification rates and the corre-

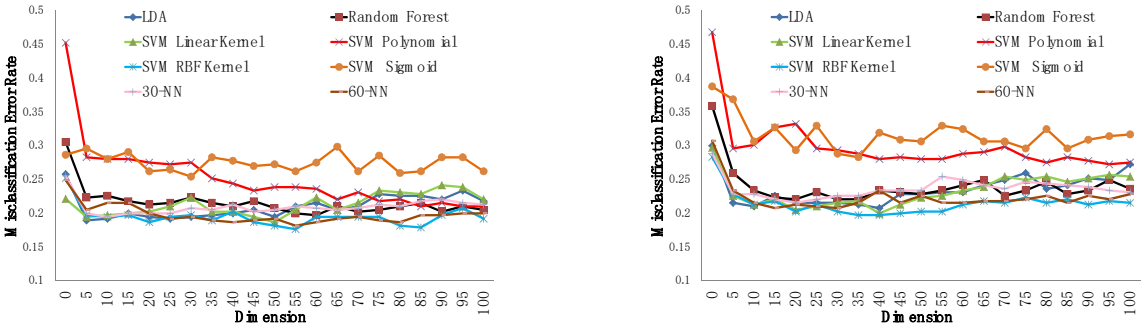


Figure 2.13: Misclassification rate as a measure of the embedding dimensions from the hippocampus via Isomap. (A) is the result of left hippocampus. (B) is the result of right hippocampus.

sponding number of dimensions from different embedding algorithms and classification algorithms (on left/right hippocampus). In that table, SVM_L indicates SVM using linear kernel (original version). SVM_R represents the radial basis kernel. RF represents the random forest approach. The first d dimensions of the embeddings are feed into the classification methods. For each different value of d the misclassification error rate L_d is calculated via a 10-fold cross-validate. Among all these L_d, the minimum one is denoted L*. And d* denotes the number of dimensions corresponding to the minimum misclassification rates L*. They are highlighted in each column of the table. The standard error for each of the entries of table 2.1 are approximately one percent. The minimum misclassification rate is around 14%, which is achieved

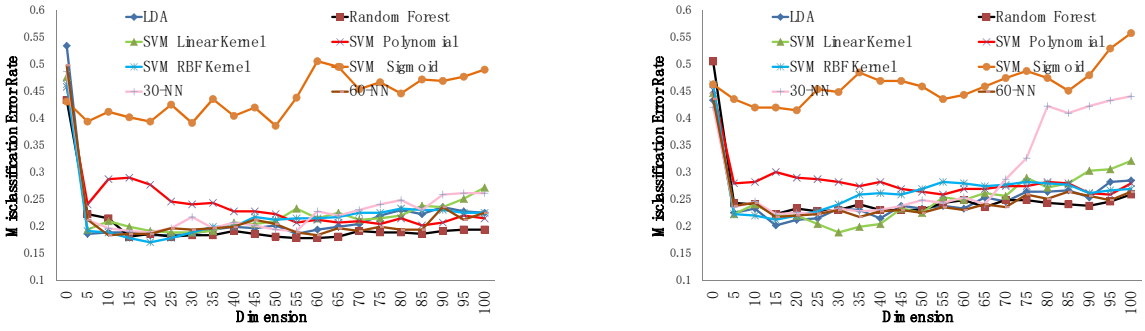


Figure 2.14: Misclassification rate as a measure of the embedding dimensions from the hippocampus via Laplacian Eigenmap. (A) is the result of left hippocampus. (B) is the result of right hippocampus.

via Laplacian Eigenmaps on the left hippocampus. Lower misclassification rates are achieved on the left hippocampus than on the right hippocampus regardless of which manifold learning algorithm is used.

Similar phenomena can be observed for the other structures. Table 2.2 summarizes the result achieved on left and right amygdala via variety of manifold learning and classification algorithms. The minimum misclassification rate is around 21% which is achieved via the Isomap method on the left amygdala. Similar to hippocampus, lower misclassification rates are achieved on the left side than on the right side for all manifold learning algorithms.

Table 2.5 is a summary to compare performance reported here with related works.

Table 2.1: The results achieved on the hippocampus. The L* denotes the minimum misclassification error rate using the first d dimensions of the embeddings calculated via different manifold learning methods. And d* is the number of dimensions corresponding to the minimum error rate L*.

Classifier	Left Hippocampus						Right Hippocampus					
	Isomap		MDS		LE		Isomap		MDS		LE	
	L*	d*	L*	d*	L*	d*	L*	d*	L*	d*	L*	d*
LDA	18.7%	9	15.3%	15	**14.0%**	**26**	20.2%	7	19.5%	25	21.3%	20
RF	19.2%	65	16.9%	16	17.9%	27	21.3%	13	21.3%	33	22.6%	6
SVM_L	18.2%	53	**14.8%**	**19**	14.8%	42	19.2%	17	**19.2%**	**27**	22.1%	3
SVM_R	**17.4%**	**53**	15.6%	15	15.6%	25	**18.9%**	**38**	20.8%	16	22.3%	3
30-NN	19.2%	24	17.1%	88	18.9%	13	21.6%	20	21.3%	55	22.3%	3
60-NN	17.7%	43	17.1%	39	18.4%	7	19.5%	26	22.1%	18	**21.0%**	**9**

Table 2.2: The results achieved on the amygdala. The L* denotes the minimum misclassification error rate using the first d dimensions of the embeddings calculated via different manifold learning methods. And d* is the number of dimensions corresponding to the minimum error rate L*.

Classifier	Left Amygdala						Right Amygdala					
	Isomap		MDS		LE		Isomap		MDS		LE	
	L*	d*	L*	d*	L*	d*	L*	d*	L*	d*	L*	d*
LDA:	21.0%	7	**18.7%**	**39**	**19.2%**	**10**	23.1%	50	21.3%	20	**21.6%**	**17**
RF:	21.6%	70	20.5%	18	21.6%	98	**22.1%**	**79**	21.0%	19	22.3%	17
SVM_L:	**20.5%**	**2**	19.7%	14	20.8%	7	23.1%	5	**20.8%**	**20**	21.6%	21
SVM_P:	22.1%	4	24.9%	7	21.0%	4	26.8%	6	28.3%	5	24.7%	57
SVM_S:	22.3%	2	25.7%	56	34.8%	11	24.9%	2	27.8%	2	37.9%	3
SVM_R:	20.8%	2	20.5%	11	20.8%	10	22.9%	27	21.0%	10	21.8%	15
30-NN:	21.3%	10	21.0%	13	21.0%	8	22.6%	14	21.6%	13	21.8%	12
60-NN:	21.0%	3	21.6%	17	20.8%	8	22.3%	32	21.0%	14	21.6%	57

Note that it only lists the methods strongly similar to our methodology, i.e., computation of dissimilarity followed by manifold learning and classification. It is difficult to compare the classification accuracy directly with other published methods, given that the difference may be caused by the different datasets used (different database or different subset of ADNI). Our embedding achieves classification accuracy comparable with related works.

Table 2.3: The results achieved on the caudate. The L* denotes the minimum misclassification error rate using the first d dimensions of the embeddings calculated via different manifold learning methods. And d* is the number of dimensions corresponding to the minimum error rate L*.

Classifier	Left Caudate						Right Caudate					
	Isomap		MDS		LE		Isomap		MDS		LE	
	L*	d*	L*	d*	L*	d*	L*	d*	L*	d*	L*	d*
LDA:	38.2%	3	32.2%	27	39.2%	6	36.9%	3	**35.6%**	**7**	39.0%	5
RF:	37.9%	36	31.7%	33	**34.8%**	**55**	41.3%	19	37.1%	5	39.0%	76
SVM_L:	37.4%	8	**30.9%**	**59**	38.2%	5	37.9%	15	37.7%	4	**37.1%**	**14**
SVM_P:	36.9%	96	36.6%	31	37.1%	55	41.8%	4	37.7%	5	39.2%	82
SVM_S:	43.9%	31	44.4%	2	42.9%	50	41.0%	2	44.9%	2	40.8%	76
SVM_R:	37.9%	86	35.3%	27	36.1%	55	37.7%	10	37.1%	70	39.0%	10
30-NN:	38.2%	32	36.1%	51	37.7%	65	**36.6%**	**25**	36.1%	78	37.9%	38
60-NN:	**36.1%**	**3**	36.6%	36	35.1%	55	37.1%	6	38.7%	77	38.4%	95

[1]The accuracy listed here is only that based on MR images (one scan for each subject).

Table 2.4: The results achieved on the ventricle. The L* denotes the minimum misclassification error rate using the first d dimensions of the embeddings calculated via different manifold learning methods. And d* is the number of dimensions corresponding to the minimum error rate L*.

Classifier	Left Ventricle						Right Ventricle					
	Isomap		MDS		LE		Isomap		MDS		LE	
	L*	d*	L*	d*	L*	d*	L*	d*	L*	d*	L*	d*
LDA:	**27.3%**	**43**	24.9%	75	**29.9%**	**87**	26.7%	32	22.9%	51	**28.4%**	**78**
RF:	31.4%	44	26.2%	58	33.2%	80	28.6%	94	24.9%	48	28.7%	77
SVM_L:	27.3%	40	24.2%	97	33.5%	87	25.6%	22	22.5%	89	31.8%	98
SVM_P:	37.1%	96	27.0%	95	35.6%	7	28.6%	73	25.4%	97	35.5%	50
SVM_S:	36.6%	56	36.9%	2	40.0%	52	31.5%	2	36.5%	41	39.8%	53
SVM_R:	31.4%	40	**21.0%**	**100**	30.6%	88	28.8%	59	**19.7%**	**81**	29.5%	92
30-NN:	35.1%	57	33.5%	71	36.1%	5	29.4%	59	28.7%	98	32.6%	26
60-NN:	34.8%	48	31.9%	39	36.1%	17	**25.3%**	**52**	29.5%	40	28.7%	58

Table 2.5: summary of representative methods in the literature. [1]

Method	[28]	[29]	[30]	[31]	[34]	Our method
{#HC}/{#AD}	25/25	69/71	37/35	116/103	57/44	210/175
Accuracy	76-84%	87%	83%	86%	77%	85%

2.4 Conclusion and Discussion

In this chapter, we presented a framework of embedding anatomical shape information into a low-dimensional space and discriminating subjects with AD from healthy controls using various classification algorithms. The LDDMM algorithm was used to measure the dissimilarity between every pair of anatomical shapes. This is in some sense a special way to extract biomarkers from MR image data. This framework achieves the comparable performance with similar application of manifold learning in discriminating subjects with cognitive dementia from healthy controls. A potential application or extension of our work is to combine the shape information extracted using this method and other imaging features extracted from different imaging modalities with their fusion improving classification accuracy [29, 30, 31].

Our result suggests that a suitable representation of anatomical shape space is inherently of low dimension. Another conclusion can be drawn from the result is that the key step in this manifold learning framework is the dissimilarity measurement because as shown in section 2.3 similar results can be achieved when different manifold learning or classification algorithms were used.

2.4.1 Comparison to Template Based Morphological Analysis

In the framework of LDDMM, we calculated a geodesic from the template coordinate space to the target coordinates. The method proposed in this paper is template-free, since there does not exist a fixed template. For template-based methods, given a fixed template, the anatomical variability within a group of subjects is encoded via the geodesic. In [11], computational tools are provided for comparing these geodesic transformations and derived a fundamental "conservation of momentum" property of these geodesics: the initial momentum encodes the geodesic connecting the template to the subject. Anatomical differences among different target groups can, therefore, be studied by analyzing the initial momentum associated with different subjects. In [47], the initial momentum space is demonstrated to be linear, and thus linear statistical analysis such as the principal component analysis (PCA) can be applied to that space. Another paper [48] successfully utilized the initial momentum space associated with volume to discriminate disease groups. In our experiment, we combined PCA and LDA to differentiate HC and AD. According to the cross-validation results, we observed, in terms of misclassification rate, 15%, 19%, 20%, and 21% respectively from the PCA+LDA procedure applied on the initial momentum space of left hippocampus, right hippocampus, left amygdala, and right amygdala. The detailed comparison of misclassification error rate is shown in table 2.6. This implies the per-

formance of our template-free method presented in this chapter is comparable to the template-based classification method.

Table 2.6: The comparison of misclassification error rate between my method (template free method) and the template based morphological analysis. The experiments are carried out on the same subset of the ADNI database. [1]

Structures	Hippocampus		Amygdala	
Sides	Left	Right	Left	Right
template based method	15%	19%	20%	21%
template free method	14.8%	18.9%	18.7%	20.8%

The advantage of template-based method is to capture subtle morphological change. Thus the subcortical structure can be further subdivided into distinct regions, on which statistical analysis can be performed. In contrast, the metric based method presented in this dissertation studies the shape space on the level of dataset. All information is encoded in a weighted graph, i.e. distance matrix. This method would benefit from increasing size of database, because in the manifold learning step, unsupervised dimension reduction methods are employed, this methodology is easily to extend to semi-supervised learning. As a result, potential improvement can be obtained from a number of MRI scans with unknown labels (no diagnosis information available).

2.4.2 Computing Time

Given a dataset of n subjects, to construct the n by n distance matrix, ${}^nP_2 = \frac{n(n-1)}{2}$ surface mappings are calculated in total. This number becomes large when n increases sharply. We utilized the power of computational resources to reduce the computing time. To be specific, we ran all these surface registration jobs on a IBM iDataPlex cluster attached to a 1 petabyte storage with 250 node and 2000 cores provided by the Institute for Computational Medicine (ICM) at Johns Hopkins University. Depending on the resolution of surface (#vertex), it takes several days to weeks to finish all surface registration jobs between every pair of surfaces. Table 2.7 shows the computing time for a surface mapping for a variety of structures using one core, i.e. one thread.

Table 2.7: Computing time for surface mapping for different subcortical structures (left side) on a single thread

Structure	#vertex	computing time (min)
amygdala	429	8
caudate	1208	55
hippocampus	1184	45
ventricle	3485	350

Chapter 3

Multiple Structure Analysis

Following the method described in chapter 2, this chapter extends the method to multiple structure analysis. The hypothesis is that different structures contain complementary information for discriminating HC and AD. A review of methods in literature associated with dissimilarity measurement and manifold learning is given in section 2.1.

The shape abnormality analysis based on metric distance measurement were explored further. Particularly a kernel based data fusion framework, i.e. Multiple Kernel Learning (MKL), was employed to combine discriminative information extracted from different structures.

In the literature, a number of methods have been proposed for computer aided Alzheimer's disease diagnosis based on metric distance measurement. A patch based analysis method was introduced in [49], where a graph kernel was computed from

structural medical images followed by kernel SVM for classification between AD patients and healthy controls. A second method also calculated dissimilarity based on image registration [50]. The standard deviation of $log(J)$ measures difference between each pair of images, where J is the Jacobian of deformation. In [51], image histogram distances were calculated to detect very mild to mild Alzheimer's disease. For both of these methods, each row of dissimilarity matrix is taken as the feature vector for classification between different cohorts.

Unlike the dissimilarity based methods above, we have multiple dissimilarity matrices. Essentially, making inferences from multiple dissimilarity matrices is a problem of data fusion. Three data fusion strategies are explored in this chapter. The most straightforward way is to simply concatenate embeddings calculated from different structures using manifold learning methods. A second strategy is to extend manifold learning method to calculate embeddings from multiple distance matrices simultaneously. The third method we explored is to transfer distance matrices into kernel matrices and make inference directly using multiple kernel learning algorithms, specifically multi-kernel SVM. In the next section these methods are introduced briefly and experimental results are presented in section 3.2. For the sake of comparison, experimental results achieved on a single structure are also presented.

3.1 Method

A first strategy is as follow first attempted. Taking the hippocampus as an example, embeddings were calculated for each data sample using manifold learning methods. Denote x_l and x_r the embeddings of a certain data sample calculated from left and right hippocampus respectively. A vector could be constructed by concatenating the first d_l dimensions of x_l and the first d_r dimensions of x_r. All possible values of $d_l = 1, \ldots, 50$ and $d_r = 1, \ldots, 50$ are attempted.

The classification method is fixed as a linear SVM. As shown in figure 3.1, the minimum misclassification error rate is 14.5% achieved at $(d_l, d_r) = (5, 18)$. As shown in section 2.3, the minimum misclassification error rate achieved on left and right hippocampi are 14.8% and 19.2% respectively. This concatenating strategy improves performance very slightly.

The second method for data fusion and inference from multiple data sources is the omnibus embedding method described in[52]. The method is named as Jointly Optimization of Fidelity and Commensurability (JOFC) and jointly optimizes fidelity of data samples. In[53], it is reported that JOFC gives a better result than SVM in a text classification problems. Given the n x n dissimilarity matrices $\Delta_1 = [\rho_{i,j}^{(}1)]$ and $\Delta_2 = [\rho_{i,j}^{(}1)]$, the $2n$ x $2n$ omnibus dissimilarity matrix M is constructed as shown in Figure 3.2. The off-diagonal block is $L = \frac{\Delta_1 + \Delta_2}{2}$. Two embeddings x_{i1} and x_{i2} can be obtained by running multidimensional scaling on M directly.

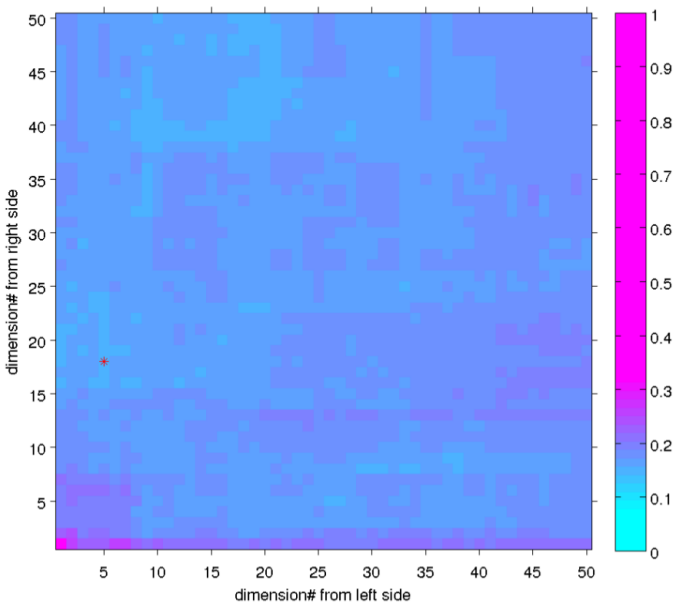


Figure 3.1: Misclassification error rate achieved by concatenating embeddings calculated from left and right hippocmpi. The minimum misclassification error is marked by a star symbol.

$$\overset{2n\times 2n}{M} = \begin{bmatrix} \overset{n\times n}{\Delta_1} & \overset{n\times n}{L} \\ L^{T} & \overset{n\times n}{\Delta_2} \end{bmatrix}$$

Figure 3.2: Omnibus dissimilarity matrix, $\Delta_1 = [\rho^{(}_{i,j}1)]$ and $\Delta_2 = [\rho^{(}_{i,j}1)]$ are two dissimilarity matrices, and off-diagonal block $L = \frac{\Delta_1+\Delta_2}{2}$.

3.1.1 Kernel Matrices Construction

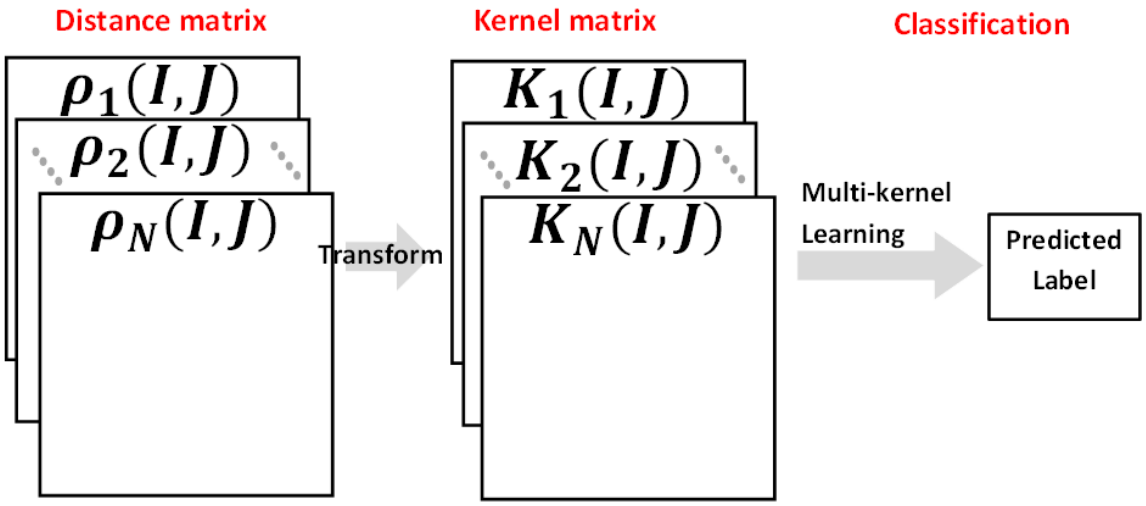


Figure 3.3: The flowchart of the framework: multi-kernel learning (Multiple kernels).

A kernel matrix is a pairwise similarity measure, i.e. a large $K(i,j)$ means data samples x_i and x_j are similar with each other. Given that a distance matrix $\Delta(i,j)$ has been calculated, there are many ways to construct a kernel matrix.

- $K(i,j) = \frac{\alpha}{\Delta(i,j)./max(\Delta(\cdot,\cdot))}$
- $K(i,j) = \alpha exp\{-\gamma(\Delta(i,j))^2\}$

The second one is used as the default way of converting a distance matrix to kernel matrix, although the first one was tested as well.

To utilize multiple kernel learning algorithms, a set of base kernel matrices $\{K^s(i,j)\}, s = 1,\ldots,S$ are constructed from distance matrices $\Delta^s(i,j), s = 1,...,S$ via formula (3.6). Here s is the index of anatomical structures. So $S = 8$ since there are four structures, hippocampus, amygdala, caudate, and ventricle on left and right sides.

Note that theoretically a kernel matrix must also be positive semi-definite. In experiments described in next section, we also tested another strategy $K(i,j) = \frac{max(D)-D(i,j)}{max(D)-min(D)}$.

3.1.2 Single Kernel SVM

Performance of classification on an individual structure is evaluated via kernel SVM. As described in section 2.2.4, kernel SVM accepts kernel matrices as input instead of original feature vectors utilizing the kernel trick. A commonly used open source SVM library, LIBSVM [43] was employed for classification between AD and HC. Since we used C-SVC there is only one hyperparamter to be determined before applying the training procedure. The trade-off parameter C as defined in section 2.2.4. A simple grid search was carried out on training data to find a reasonable value for C.

In contrast to manifold learning methods described in chapter 1, this kernel based method does not need to determine how many dimensions of embeddings to feed into classification.

3.1.3 Multiple Kernel SVM

The Multiple Kernel Learning (MKL) method seeks an optimal combination of multiple kernels obtained from different data sources [54, 55]. The combination could

be linear or nonlinear. Here we only consider linear combination. It is an extension of single kernel SVM which is described in section 2.2.4. In the Multi-Kernel SVM (MK-SVM) framework, we have S kernel Matrices. Let $K^{(k)}$ be the k^{th} kernel. MK-SVM seeks a set of linear mixture coefficient, $\beta \in R^S$:

$$K(x_i, x_j) = \sum_{k=1}^{S} \beta_k K^{(k)}(x_i, x_j) \tag{3.1}$$

In section 2.2.4, the dual problem is defined as the one in which the objective function only depends on the kernel matrix of the dataset. This combined kernel K was expected to give better results on test data compared with results from each individual kernel $K^{(k)}$. The most straightforward way to find the optimal mixture coefficient is searching over the whole space. This is not practical when the number of kernels is large. Thus MK-SVM was proposed to optimize over α and β simultaneously.

$$\begin{aligned} \mathcal{S}(\alpha) = & \quad \frac{1}{2}\alpha^T Q \alpha - e^T \alpha \\ = & \quad \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} y_i y_j \alpha_i \alpha_j K(x_i, x_j) \end{aligned} \tag{3.2}$$

The objective function of the dual problem can be written as 3.2. $S(\alpha)$ can be rewritten as a function of α and β by combining 3.1 and 3.2. The dual problem of MK-SVM can be written in a formal way:

$$\min_{\beta} \max_{\alpha} S(\alpha, \beta) \tag{3.3}$$

$$with \sum_{i=1}^{m} y_i \alpha_i = 0, 0 \leq \alpha \leq C1, \tag{3.4}$$

$$\beta \geq 0, \|\beta\|_1 = 1 \tag{3.5}$$

In our pipeline, a Multi-kernel SVM with p-norm regularization using the Sequential Minimal Optimization (SMO) [56] was employed to train and test on multiple kernel matrices transformed from distance matrices calculated using LDDMM. The objective in MKL is to jointly learn kernel $K = \sum_i d_i K_i$ and SVM parameters w, b, ξ from training data $\{(x_i, y_i)\}$.

$$\begin{aligned} &\min_{w,b,\xi \geq 0, d \geq 0} \frac{1}{2} \sum_k w_k^t w_k + C \sum_i \xi_i + \frac{\lambda}{2} (\sum_k d_k^p)^{\frac{2}{p}}, \\ &s.t. y_i (\sum_k \sqrt{d_k} w_k^t \phi_k(x_i) + b) \geq 1 - \xi_i \end{aligned} \tag{3.6}$$

The third term in the objective function is the regularization on kernel weights. This term is necessary because d_k could be arbitrary large. A large enough d_k ensures that constraining $y_i(\sum_k \sqrt{d_k} w_k^t \phi_k(x_i) + b) \geq 1 - \xi_i$ always holds. We followed the optimization method proposed in [56], where the p-norm squared regulariser was limited with $p > 1$.

3.2 Experiments and Result

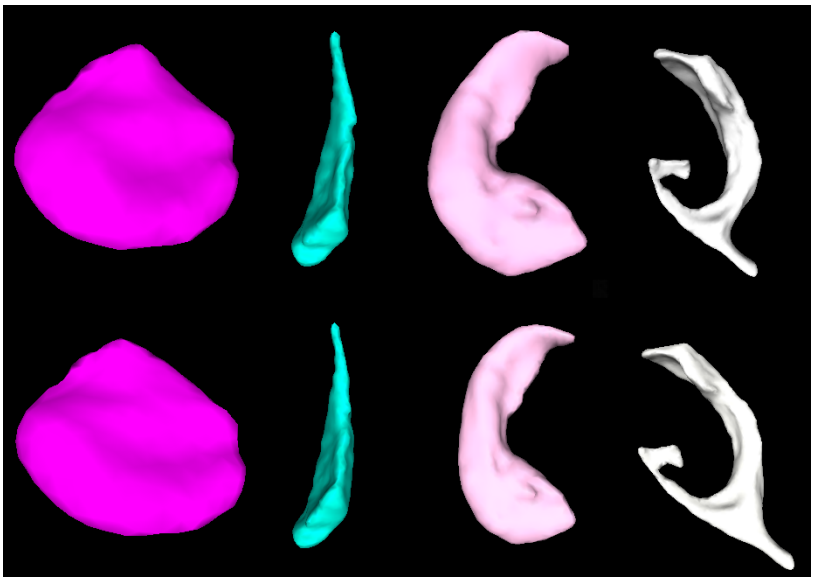


Figure 3.4: An example of iso-surfaces of four structures, from left to right: amygdala, caudate, hippocampus, ventricle on left side (top row) and right side (bottom row).

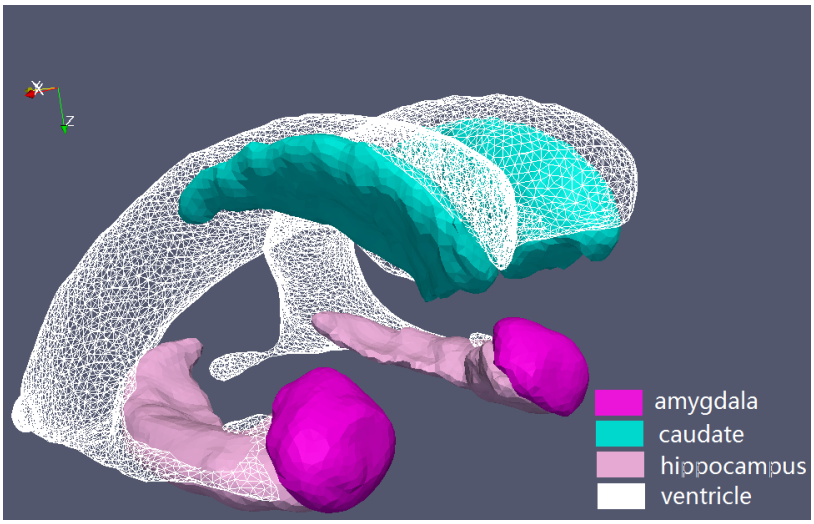


Figure 3.5: Some examples of iso-surfaces of Hippocampi with AD in red and HC in green.

The iso-surfaces contouring subcortical structures were extracted as described in

section 1.3.1. Figure 3.4 visualizes four structures on both sides. Figure 3.5 demonstrates the relative position between these structures.

Chapter 2 demonstrated the result of classification using embeddings calculated via manifold learning algorithms. A phenomenon is the asymmetry between structures on left and right sides. For example, using MDS and LDA in manifold learning and classification step, the minimum misclassification error rate is 15.3% and 19.5% on left and right hippocampus. Similar results can be observed from other structures. No matter what manifold learning and classification methods are employed, a better performance is achieved with the left side. As described in section 3.1, embeddings calculated from left and right hippocampi are concatenated to form a longer feature vector. This simple method could provide a baseline of combined information extracted from different structures. We compare the minimum misclassification error rate achieved on each side by combining both sides. Taking hippocampi as an example, we fixed the linear SVM as the classification algorithm to compare whether a lower misclassification error rate can be obtained when concatenating the first d_l dimensions and the first d_r dimensions from embeddings calculated from the left and right sides respectively.

As table 3.1 shows, only slightly improvement is observed when combining the left and right hippocampi using the concatenating strategy.

Similar results can be observed on amygdalae presented in table 3.2. In other words, hippocampi or amygdalae on two sides are highly correlated with each other.

Table 3.1: The results achieved on the hippocampus using the concatenating strategy. The L* denotes the minimum misclassification error rate using the first d_l dimensions of the embeddings calculated from left hippocampus and the first d_r dimensions of the embeddings calculated from right hippocampus. And (d_l^*,d_r^*) are the numbers of dimensions of the embeddings calculated from left and right hippocampus corresponding to L*

Manifold Learning	Left	Right	L*	d_l^*	d_r^*
MDS	14.8%	19.2%	14.5%	5	18
Isomap	17.4%	18.9%	16.9%	17	11
LE	14.8%	22.1%	14.7%	27	13

This simple strategy of data fusion is not efficient.

Table 3.2: The results achieved on the amygdala using the concatenating strategy. The L* denotes the minimum misclassification error rate using the first d_l dimensions of the embeddings calculated from left amygdala and the first d_r dimensions of the embeddings calculated from right amygdala. And (d_l^*,d_r^*) are the numbers of dimensions of the embeddings calculated from left and right amygdala corresponding to L*

Manifold Learning	Left	Right	L*	d_l^*	d_r^*
MDS	19.7%	20.8%	19.1%	19	7
Isomap	20.5%	23.1%	20.0%	9	3
LE	20.8%	21.6%	20.4%	10	4

The second strategy, i.e. JOFC was employed to combine discriminative information extracted from the same structure on different sides. One issue is to determine the dimensionality. As usual the first d dimensions which dominate 95% of the variance are kept. For convenience of comparison, first d dimensions were selected and used for classification, where $d = 1, \ldots, 100$. For each different value of d, a misclassification error rate L_d can be calculated via a 10-fold cross-validate. The minimum

misclassification error rate is shown in table 3.3 (column "L^*") along with d^*, the corresponding the number of dimensions. To compare the result achieved on one single structure, L_l^* and d_l^* indicate lowest error rate and corresponding number of dimensions on left side, while L_r^* and d_r^* are those on right side.

Table 3.3: The results achieved via JOFC and comparison with results achieved on left or right structure alone. The L^* denotes the minimum misclassification error rate and the d^* denotes the corresponding number of dimensions. The (L_l^*, d_l^*) and (L_r^*, d_r^*) are results achieved on left and right structures alone. The classification algorithm is fixed to linear SVM.

Structure	L_l^*	d_l^*	L_l^*	d_l^*	L*	d^*
Amygdala	19.7%	14	20.8%	20	19.9%	21
Caudate	30.9%	59	37.7%	4	29.1%	38
Hippocampus	14.8%	19	19.2%	27	15.2%	16
Ventricle	24.2%	97	22.5%	89	22.6%	43

One can observe from table 3.3, the L^* increases slightly when combining left and right sides via JOFC. The only exception is caudate, on which combination improves the performance. One benefit from JOFC is that the d^* decrease compared with d_l^* and d_r^* in general.

Chapter 2 showed that classification performance varies for different dimensionality. The optimal dimensionality corresponding to the minimum misclassification error rate was reported. But in practice, we cannot determine the optimal number of dimensions in advance. A more practical way is to use kernel-based methods, e.g. kernel SVM since the dissimilarity measurement can be transformed into kernel matrices easily. An example of distance matrix and the corresponding kernel matrix is

shown in figure 3.6.

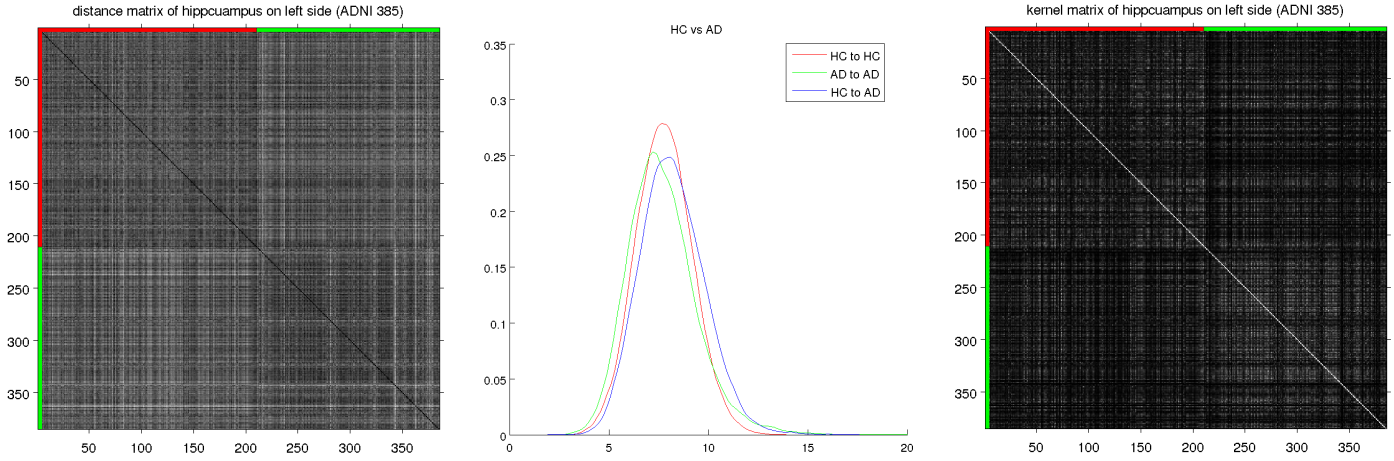


Figure 3.6: Distance matrix calculated on left hippocampus and generated kernel matrix. (A) Distance matrix of left hippocampus. (B) Empirical distribution of distance within and across different groups. (C) Kernel matrix generated using formula $K(i,j) = \alpha exp\{-\gamma(D(i,j))^2\}$. Here $\sigma = 0.05$.

In order to improve classification accuracy, the multi-kernel learning method as described in section 3.1.3 is employed to combine information extracted from different structures. Two approaches were tried for converting distance matrices to kernel matrices.

- Approach 1: $K(i,j) = \frac{\alpha}{D(i,j)./max(D)}$
- Approach 2: $K(i,j) = \alpha exp\{-\gamma(D(i,j))^2\}$

Table 3.4 shows misclassification error rate obtained using single-kernel SVM and multi-kernel SVM. The first approach of converting distance matrices into kernel matrices was employed for kernel matrix construction, i.e. $K(i,j) = \frac{\alpha}{D(i,j)./max(D)}$. The bottom row is the normalized weights d_i for different kernels learned by MK-SVM

introduced in section 3.1.3. Note that the distance matrices were normalized into range [0,1] before feeding them into MK-SVM. This ensures that the multiple kernels also have the same range and thus the normalized weights is meaningful to evaluate the importance of different structures. Another observation is a higher classification performance of MK-SVM (column "MK-SVM") than the single-kernel SVM on any structure (column "SK-SVM"). This implies different cortical structures contain complementary information and multiple kernel SVM is effective in combining discriminating information extracted from different cortical structures. The weight for hippocampus is higher than other structures, which provides evidence that hippocampus contains more discriminative information. Besides, asymmetric performance was observed on every structure. In each round of cross-validation, grid search was carried out for hyper-parameter selection. In general, a typical setting for this problem is $\sigma = 0.15$ for kernel matrix construction, $p = 3.5, C = 7.0$ in multiple kernel SVM. The weights shown in table 3.4 are averaged among 5 rounds of cross-validation.

Table 3.4: Misclassification error rate (MCR) achieved on different structures via Single-Kernel SVM (SK-SVM) and Multi-Kernel SVM (MK-SVM) on all structures using approach 1 for kernel construction. The bottom row lists normalized weights of each kernel learned in MK-SVM.

	SK-SVM								MK-SVM
	Hippocampus		Amygdala		Caudate		Ventricle		All
	Left	Right	Left	Right	Left	Right	Left	Right	
MCR	16.56%	19.24%	19.48%	21.70%	29.61%	35.34%	26.17%	28.81%	14.04%
Weights	0.2042	0.1887	0.1757	0.1654	0.0564	0.0569	0.0827	0.0698	NA

Similar results can be obtained using approach 2 for kernel construction, i.e. $K(i,j) = \alpha exp\{-\gamma(D(i,j))^2\}$. Table 3.5 shows misclassification error rate using this approach. The bottom row lists normalized weights of each kernel learned in MK-SVM. Similar as table 3.4, this implies different structures contain complementary information and multiple kernel SVM is effective for combining discriminating information extracted from those structures. The highest weight is assigned to left hippocampus, which validates that left hippocampus contains more discriminative information. A typical setting for this approach is $\alpha = 0.05$ for kernel matrix construction, and $p = 2.5, C = 10.0$ for MK-SVM.

Table 3.5: MCR achieved on different structures via SK-SVM and MK-SVM on all structures using approach 2 for kernel construction.

	SK-SVM								MK-SVM
	Hippocampus		Amygdala		Caudate		Ventricle		All
	Left	Right	Left	Right	Left	Right	Left	Right	
MCR	16.14%	20.03%	20.02%	21.81%	30.89%	35.59%	28.07%	29.98%	13.83%
Weight	0.2155	0.1695	0.1567	0.1718	0.0588	0.0490	0.0908	0.0880	NA

Figure 3.7 compared combination coefficients learned from MK-SVM. Different colors indicate different approaches for kernel matrix construction. In principle, higher weights were assigned to hippocampus and amygdala individually. While the caudate and ventricle have weights relatively low. This makes sense since higher performance is achieved on the hippocampus and amygdala.

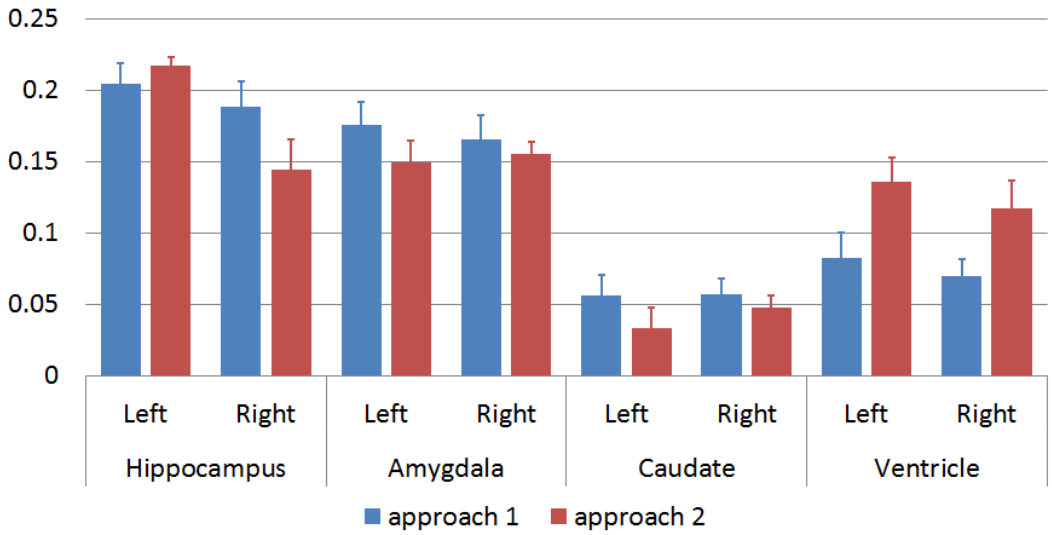


Figure 3.7: Mean value and standard deviation of Combination coefficients learned from multiple kernel SVM.

3.3 Conclusion and Discussion

In this chapter, we present a framework of multiple structure analysis for discriminating subjects with AD from healthy controls using kernel based classification algorithms. Kernel matrices are calculated from dissimilarity measurements of multiple cortical structures via the LDDMM surface registration algorithm. The result demonstrates multi-kernel learning algorithm is effective for combining complementary information extracted from multiple cortical structures. The performance of discriminating subjects with cognitive dementia from healthy controls achieved in this chapter is comparable with related work, but only based on some regions instead of the whole brain as in related work.

Some potential extension of the works presented in this chapter include using in-

tensity information in the dissimilarity measurement in addition to shape information, and combination of morphological information in form of pairwise dissimilarity.

Chapter 4

Fusion of Geometric Information and Functional Contrast

Recently structural images, i.e. MRI scans, are increasingly used to support computer aided Alzheimer's disease diagnosis besides the traditional neuropsychological examination. In this chapter, our discriminative analysis between AD patients and healthy controls extends from vertex-based method to voxel-based method. In other words, 3D T1-weighted MR images are analyzed instead of 2D surfaces as in previous chapters.

As shown in section 1.6, in literature, three categories of related methods have been proposed according to the type of features extracted from MR images for classification. 1) volumetric measurements of regions of interest (ROI), 2) vertex-based methods, and 3) voxel-wise features. The method introduced in this chapter extracts voxel-wise

geometric and photometric features from seven structures. The analysis was carried out on ROIs, i.e., subcortical structures. But features were extracted at voxel-level from structural images. Specifically we focused on seven subcortical structures (on both left and right sides), amygdala, caudate, hippocampus, pallidum, putamen, thalamus, and ventricle.

Inspired by the of deformable template models[13], which were firstly proposed in the computer vision community, we explored the idea of separating shape information and appearance information through nonlinear registration. In[13], the aim is to compare two images, or equivalently to calculate a distance between two images. In that model the difference between images comes from two parts, geometric deformation, and intensity shift. We borrow this general idea in our application. But instead of calculating a metric distance, we directly extract the raw feature, i.e. the parameters of maps calculated via a non-linear image registration algorithm and the intensity value of each voxel after registration. The assumption is these two kinds of features are complementary, thus higher performance of classification can be expected when combining them together.

On the other hand, some recent works[57] have shown that the intensity value of structural images can be used to discriminate between AD and HC, as well as MCI and HC. However, most of them used the average intensity of some structures. In photometry framework, we use the intensity value as functional contrast feature on voxel level.

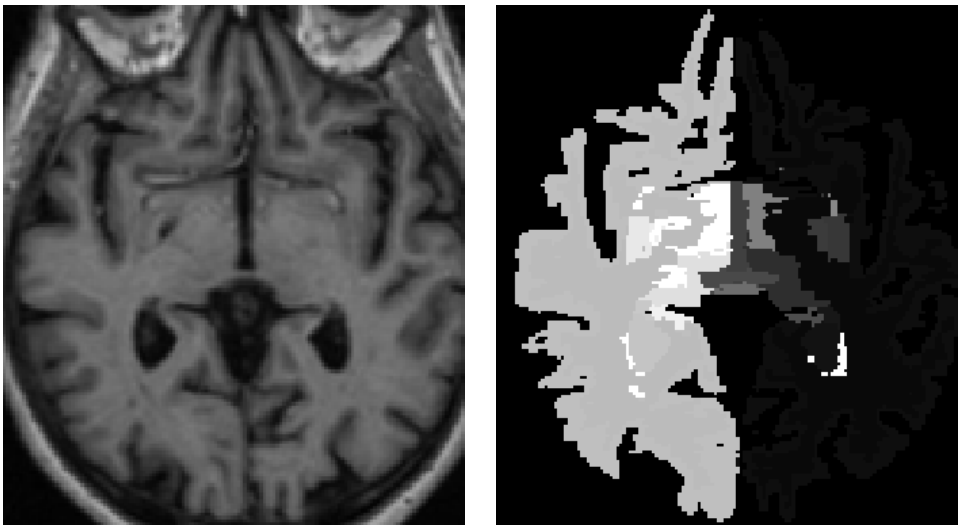
From the perspective of neuroanatomical analysis, the method here can be considered as a combination of Deformation-Based Morphometry (DBM) [58] and Voxel-Based Morphometry (VBM) [59]. DBM methods compare neuroanatomy of human brains in terms of parameterized deformation fields. It analyzes deformation fields that map individual brain to a standard template. In contrast, VBM compare brain images on a voxel-level basis after deformation been applied for spatial normalization. In this chapter, geometric features represent that extracted from the deformation fields. And photometric features or functional contrast indicate the intensity value of voxels after alignment.

In section 4.1 the method is described in detail. Experiments and results are presented in section 4.2. This chapter ends by a section to summarize the conclusions and future work.

4.1 Method

Our pipeline of deformable template model starts from segmentation as described in chapter 1. For the i-th subject, we have two images, gray-value T1 weighted MR images T_i and the corresponding segmentation images J_i ($i = 1, \ldots, N$). The segmentation image J_i has limited gray levels with different gray value indicates different structures in the human brain. In this dissertation it is considered as a set of binary images, i.e., $J_i = \{J_i^{(s)}\}$, where $s = 1, \ldots, M$ is the index of structures. Each

of these $J_i^{(s)}$ defines the contour of a certain structure for the i-th subject. Figure 4.1 shows an example of a T1 image T_i and the corresponding segmentation J_i.



(A) (B)

Figure 4.1: Example of a T1 image (A) and the corresponding segmentation (B).

Given T1-weighted images and segmentation images, the photometric features and geometric features are extracted via a pipeline as shown in figure 4.2. The segmentation images are Registered to the template segmentation images via multi-channel LDDMM. The transformations are then applied to the T1-weighted images, so that all images are transformed into a common coordinate system. Geometric features can be extracted from the parameterized transformation $\phi_1(x)$. For example the norm of the deformation vector field $\|\phi_1(x)\|_{l-2}$. Section 4.1.4 shows more details about this. The intensity value of each voxel of transformed target image forms the photometric feature or the functional contrast. For the high dimensionality of the raw features, an optional step following the feature extraction is feature selection or

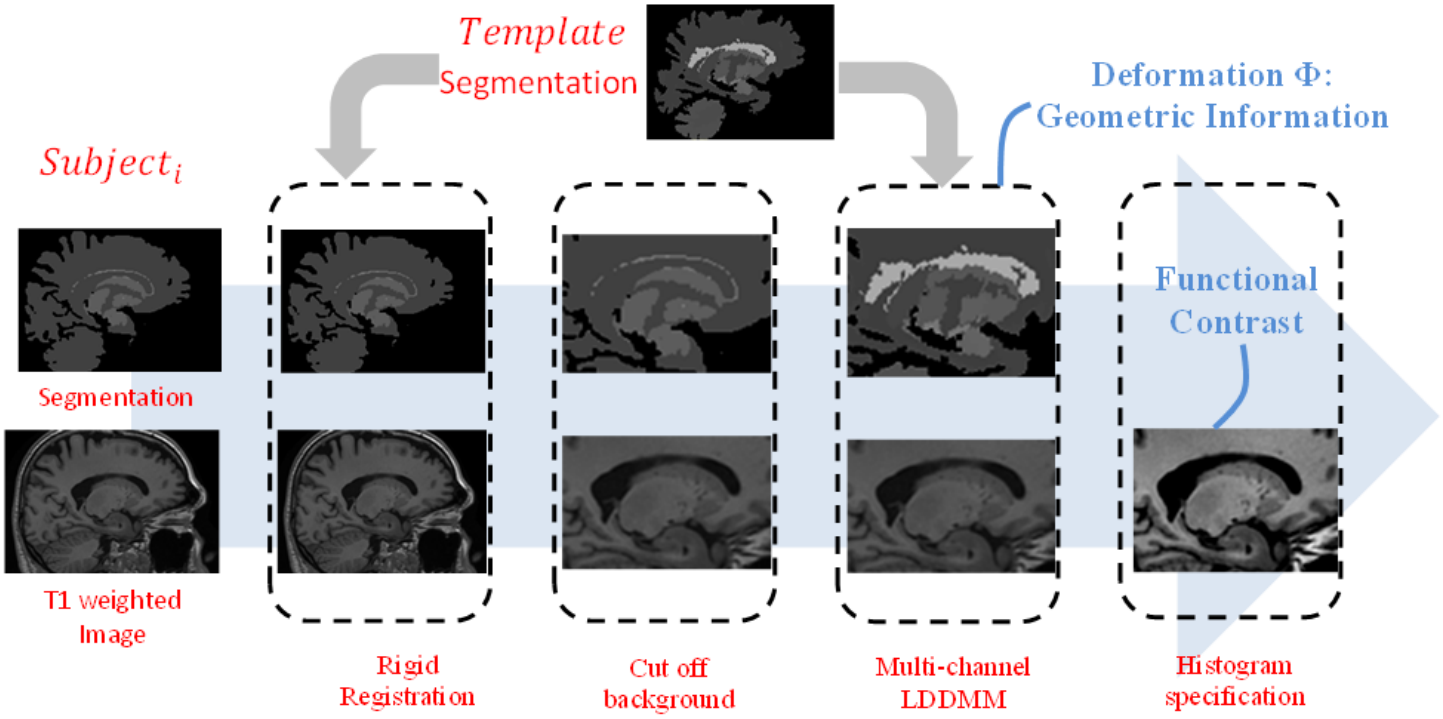


Figure 4.2: The flowchart of feature extraction component.

dimension reduction. Another optional step is the data fusion because two feature vectors are extracted from different sources. The last step is to train classifiers and make prediction for testing data. The overall flowchart is shown in figure 4.3. In the rest of this section the method is described in detail.

4.1.1 Multi-Channel LDDMM

In order to capture the shape variation of subcortical structures, a deformation from the segmentation images of each subject to a common target segmentation image need to be calculated. This is a necessary step before photometric feature extraction, because intensity value is meaningless without alignment.

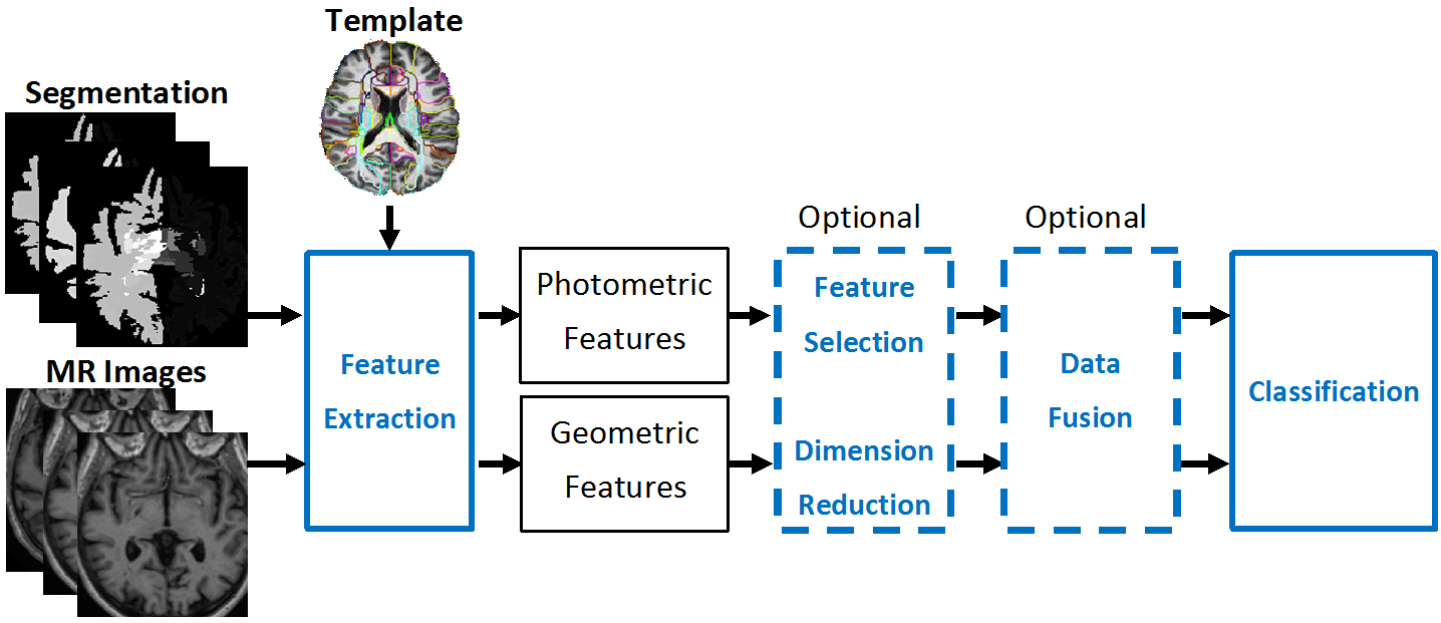


Figure 4.3: The flowchart of our method.

As a preceding step affine registration was carried out to remove the variation caused by different poses in image acquisition procedure. An open source library, Automated Image Registration (AIR) [60, 61, 62], was employed to calculate affine transformations. The performance of this affine registration is critical since if template and target image does not match with each other well then in the following non-linear deformation registration (LDDMM) step, the calculated mapping encodes much irrelevant information which dominant over geometric discriminant information.

The following step is non-linear registration. Multi-channel LDDMM (MC-LDDMM) method[63] is used to register all structures simultaneously. There are several reasons to choose this approach. First, it is closer to real world, i.e. all structures grow or develop together. So it is more likely to capture real geometric variation. Second, it provides some flexibility to register all structures simultaneously for we could adjust

weights of different structures in the matching problem. The objective function of MC-LDDMM is very similar to single contrast version as described in section 1.4. Formally given two sets of images, template and target, denoted $I_{temp} = \{I_{temp}^{(s)}\}$ and $J_i = \{J_i^{(s)}\}$, where $s = 1, \ldots, M$ and M is the number of channels. In our case $M = 14$ is the number of structures. MC-LDMM seeks the optimal time dependent velocity field v_t^* to minimize energy function composed by two parts as shown in equation 4.1. A diffeomorphic mapping $\phi_1(x)$ can be generated by integrating the smooth optimal velocity field $v_t^*(x)$ with initial condition being identity $\phi_0(x) = x$. Different from single contrast LDDMM, the mismatching error is the sum of mismatching error of every channel. Besides trade-off between smoothness and matching, the $\sigma_s, s = 1, \ldots, M$ can be used to adjust weights between different channels.

$$v_t^* = \arg\min_{v \in V} \int_0^1 \|v_t\|_v dt + \sum_{s=1}^{M} \frac{1}{\sigma_s^2} \|I_{temp}^{(s)} \circ \phi_1^{-1} - J_i^{(s)}\|_{L^2} \tag{4.1}$$

In the energy function above $\|v_t\|_v = \|Lv_t\|_2^2 =< L^\dagger L v_t, v_t >_{l^2}$ is a norm defined by a operator L as stated in section 1.4. The operator $L = Id + \alpha\nabla^2$, where Id is identity, and ∇^2 is the Laplacian operator which enforce some smoothness of the velocity field and thus guarantee the ϕ_1 is a diffeomorphism [14]. The varying parameter α controls the smoothness of the velocity field and deformation. A smoother deformation is generated with a larger α. And with a smaller α, a more accurate transformation is calculated. In order to register images from

Table 4.1: Segmented structures and σ_c in MC-LDDMM

Structure	weighting
Amygdala	$\sigma_1 = 0.8$
Caudate	$\sigma_2 = 0.75$
Hippocampus	$\sigma_3 = 0.5$
Pallidum	$\sigma_4 = 0.8$
Putamen	$\sigma_5 = 1$
thalamus	$\sigma_6 = 1$
ventricle	$\sigma_7 = 1$

rough to fine, a cascading α MC-LDDMM method was used with several iterations. In each iteration, MC-LDDMM is employed with α varies from large to small. Specifically our registration is a 9-iteration procedure, where the α list is $\{0.05, 0.025, 0.01, 0.005, 0.0025, 0.001, 0.0005, 0.0001, 0.00005\}$. Table 4.1 shows the weights we used for different structures in MC-LDDMM.

4.1.2 Local Histogram Specification

After spatial normalization, i.e. registration, all images were transformed into a common template coordinate system. Another preprocessing step, histogram specification, is necessary to eliminate illumination variations. It aims at adjusting intensity value of images to have the same distribution or histogram. In order to accomplish better result, this is carried out on each subcortical structure individually. The reason is that the volume of subcortical structures may vary in a population. Especially the varying size of ventricles would affect the results of histogram matching.

4.1.3 Feature extraction

For the i-th subject, the transformation $\phi_1^{(i)}(x)$ mapping the segmentation image J_i to the template segmentation image I_{temp} has been calculated via MC-LDDMM. Similar with some previous work [64, 65, 66, 67], our method also considers a compacted representation of the geometric information, i.e. the $l-2$norm of the transformation $N^{(i)}(x) = \|\phi_1^{(i)}(x)\|_{l-2}^2$ or another scalar field $\mathcal{J}^{(i)}(x) = log(det(D\phi_1^{(i)}(x))$, where $D\phi_1^{(i)}(x)$ is the Jacobian matrix of $\phi_1^{(i)}(x)$. This scalar field $\mathcal{J}^{(i)}(x)$ measures the expansion or atrophy at each voxel. Positive values indicates expansion, while negative values denote atrophy of a region relative to the template. The geometric feature vector of the i-th subject, $F_g^{(i)}$ is formed by stacking the $N^{(i)}(x)$ or $\mathcal{J}^{(i)}(x)$ into a vector. The T1-weighted image $T_i(x)$ is transformed towards the template segmentation. Denote $\hat{T}_i = T_i \circ \phi_1^{(i)}$ the transformed T1-weighted image. The photometric feature vector of the i-th subject, $F_p^{(i)}$ is formed by stacking the $\hat{T}_i(x)$ into a vector.

4.1.4 Feature selection and dimension reduction

The original raw photometric and geometric features were obtained after the steps as described above. Although some classification algorithms can handle the high dimensionality of the raw features, it is still worthwhile to seek a better representation of these raw features. Two feature selection technologies are explored, i.e., **T-test** based method and **random forest** based method. For dimension reduction, the

principal component analysis (PCA) is employed.

Feature selection is the process of selecting a subset of features which is relevant for training models. This technique is often used in domains where dimensionality of features is much larger than the number of training samples. In our case the dimension is over 80,000. While the number of training data sample is merely 385. Two feature selection methods were tried in our project.

- **T-test**: The t-test or student's test is a statistical hypothesis test to compare two normally distributed samples or populations. The null hypothesis is that the means of two sets of data samples are equal. When the p-value of the t-test is smaller than a threshold like 0.05, the null hypothesis is rejected. This means the two sets of samples are of different means. The dimensions with p-value <0.05 were kept. The dimensionality of feature vector decreases from over 80,000 to 7,000 approximately via this simple feature selection process. To have a more stable performance the corresponding non-parametric statistical hypothesis test, Wilcoxon signed-rank test was also tried.

- **Random Forest**: Random forest is a classification or regression algorithm applying the idea of bootstrap aggregating (bagging) to decision tree. Besides classification or regression, another very important usage of random forests is to calculate the importance of each variable in feature vectors [68, 69]. A variable importance is calculated during the training procedure which is measured through permutation. For each decision tree a number of data samples are ran-

domly selected to train the data. The remaining data samples are called data samples out-of-bag (OOB). A certain variable is randomly permutated in the OOB samples. The increase in mean of the error of the decision tree is used to measure the importance of the variable. All variables are sorted in decreasing order of RF variable importance. The last step is to compare OOB classification error when using first k variables. Here $k = \{5\%, 10\%, \ldots, 95\%, 100\%\} * N$, where N is the number of variables in original feature vectors.

Dimension reduction is the process of transforming the original features into a set of new features, which as usual is in a lower dimensional space. In contrast to feature selection, feature extraction transforms original feature vector into another representation. The transformation could be linear or nonlinear. The most commonly used feature extraction or dimension reduction technique is principal component analysis (PCA). The original input feature vectors are projected onto some directions (principal components) which corresponding to largest variances. In practice k is chosen such that the first k principal components that account for 95% of the total variance. There is another method for dimension-reduction, i.e. independent component analysis (ICA) [70]. This method can be an alternative to PCA.

4.1.5 Classification and Data Fusion

In the classification step, PCA+LDA and linear SVM as described in previous chapters were employed. For each subject, there are two feature vectors, photometric

and geometric feature vectors. we could split the feature vectors furthermore into feature vectors extracted from different structures, e.g. the photometric feature vector extracted from the left hippocampus. Thus in total there are 28 feature vectors for each subject. To combine information from different sources, data fusion techniques were used. According to the level on which combination occurs, data fusion methods can be categorized into three groups. 1) Low-level combination. That is to concatenate feature vectors to form a longer feature vector. This is the simplest way for data fusion. 2) Medium-level combination. This category of methods extracts medium-level features from different sources and concatenates them together to feed into classifiers. 3) High-level combination. The methods fall into this category make a prediction based on the predictions based on information extracted from different sources individually. In some contexts, it is named label fusion. Three data fusion technologies are explored here. To be specified, they are **majority voting**, **two-level SVM**, and **multi-kernel SVM**.

Firstly, the majority voting is carried out to make the overall prediction. This is a meta-method of ensemble learning. In our case, a SVM classifier was trained on each structures and each kind of features (photometric or geometric). The label is predicted as the one most classifiers vote. Specifically we have twenty-eight classifiers to vote.

Secondly, the two level SVM [71] method was employed to extract medium-level features. On each feature space, an ordinary SVM classifier (low-level SVM) is

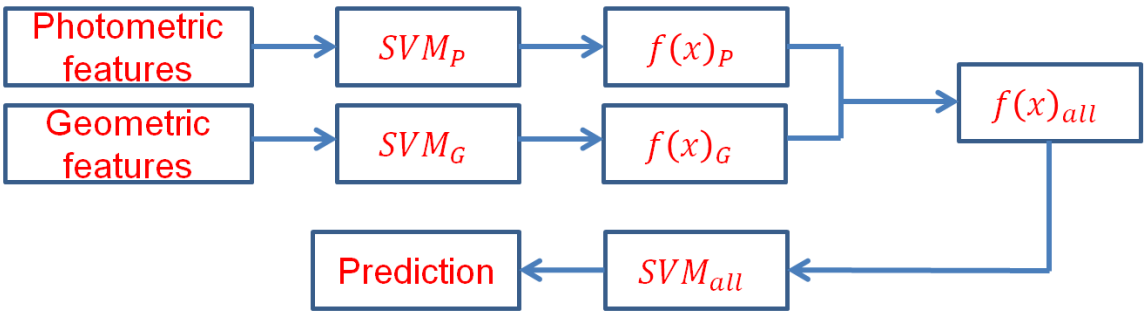


Figure 4.4: Flowchart of two-level SVM to combine photometric and geometric features.

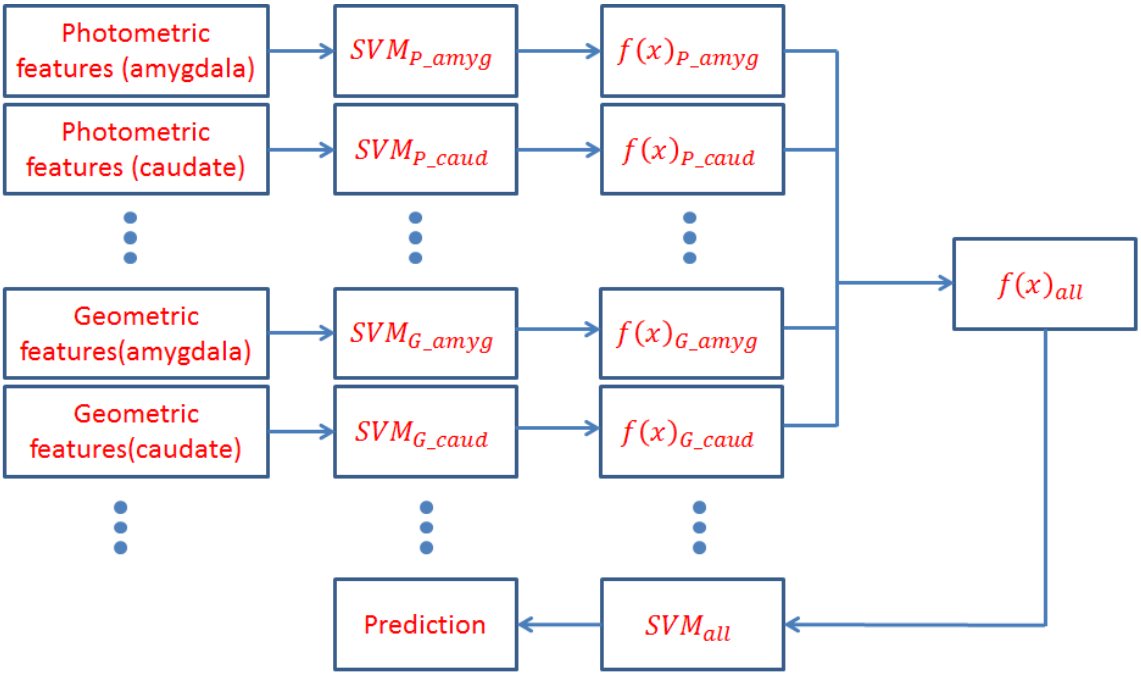


Figure 4.5: Flowchart of two-level SVM to combine features extracted from different structures.

trained. Then the output of the prediction function is a scalar feature. Each data source corresponds to a scalar feature. Thus a feature vector is formed on which a high-level SVM is trained. There are two different ways to define data sources in our problem. One way is to train low-level SVMs on geometric features and photometric features individually. Then a high-level SVM is trained in the two dimensional space. The other one is to train low-level SVMs on two kinds of features extracted from different structures individually. We have seven structures on both sides. In total, the dimensionality is twenty-eight for high-level SVMs. These two approaches are demonstrated in figure 4.4 and 4.5 respectively.

Finally, the kernel based data fusion method was also tested [72]. As discussed in section 3.1.3, multi-kernel SVM (MK-SVM) seeks a set of linear mixture coefficients to combine kernel matrices into one. This combined kernel matrix can be used to train classifier using kernel based method, e.g. SVM. A popular kernel function, the (Gaussian) radial basis function kernel, or RBF kernel, was employed to calculate kernel matrices on different structures and different features. MK-SVM are used to combine these kernels together and make a overall prediction.

4.2 Experiments and Result

We validated the proposed methodology by classifying AD and healthy controls on a large MR image database, ADNI as described in section 1.3. Note that an

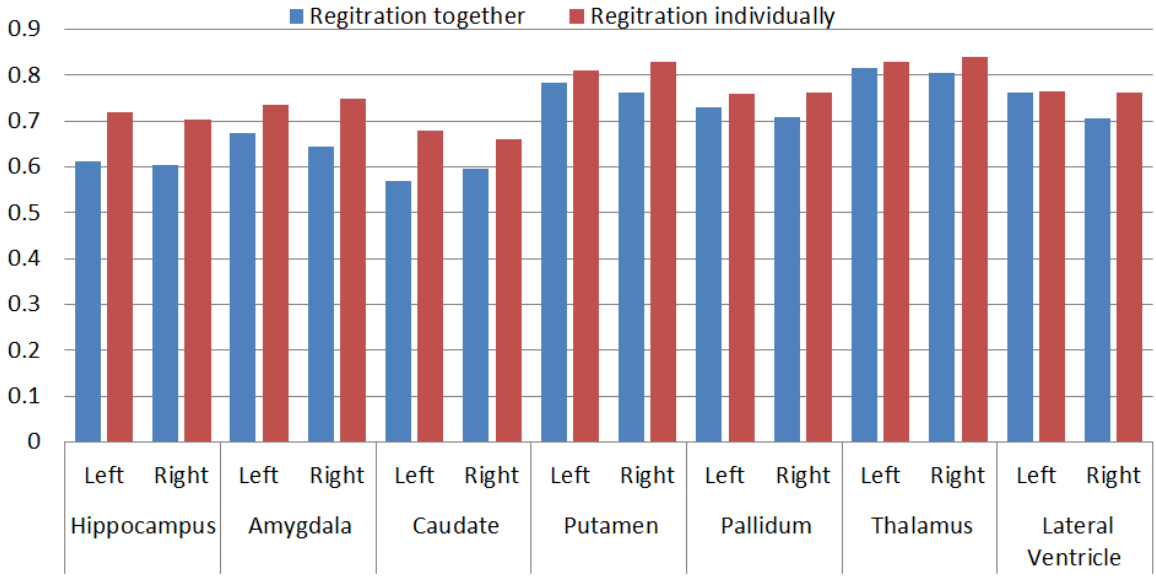


Figure 4.6: Rigid registration result (Dice's score).

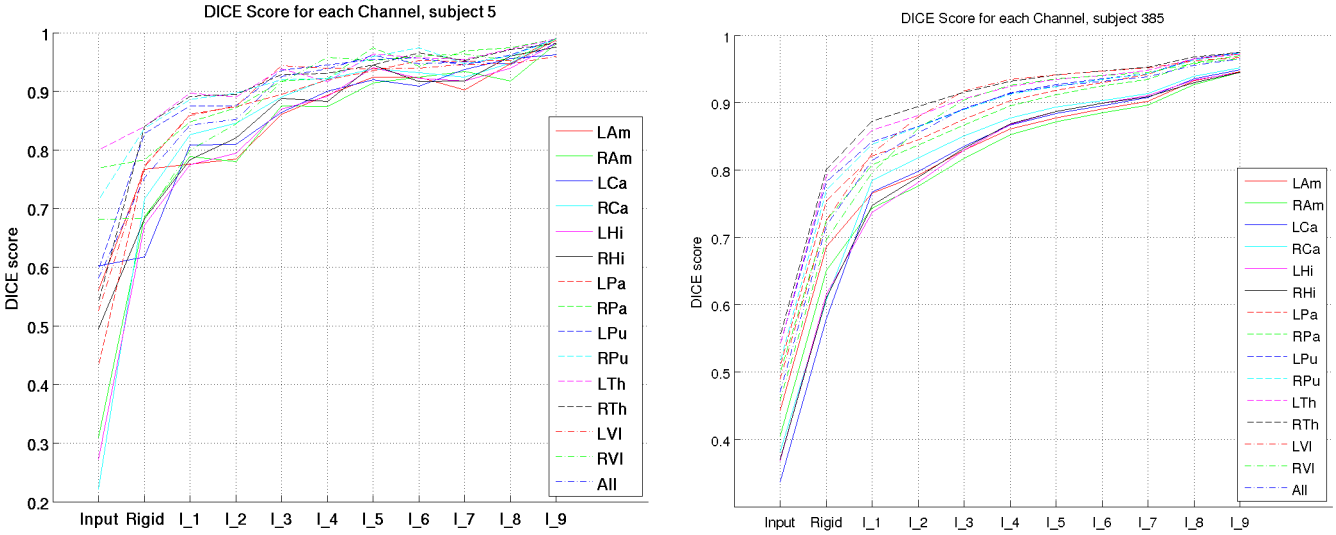
atlas was used as the target segmentation image. All T1 weighted MR images were transformed towards this common target segmentation image.

4.2.1 Image Registration

An affine registration is firstly carried out via AIR. To evaluate the performance of affine registration, the Dice's coefficient was calculated, which is defined by $DC = \frac{2\cdot\|A\cap B\|}{\|A\|+\|B\|}$, where A and B are template and target images (binary images) and $\|\cdot\|$ is the cardinality of a set. Figure 4.6 compares the Dice's coefficient after rigid registration of all structures together and individually. One can observe that template and target segmentation have large portions which overlap (>0.6 on average).

Then a 9-round MC-LDDMM was carried out for non-linear registration. In fig-

ure 4.7 (A) shows DICE score for a certain subject (the 5th subject in our database) and (B) is the averaged DICE score over all subjects. Note that the I_i on x-axis indicates the i-th iteration in multiple round LDDMM. The legends {LAm, RAm, LCa, RCa, LHi, RHi, LPa, RPa, Lpu, RPu, LTh, RTh, LVl, RVl, All} indicate {left amygdala, right amygdala, left caudate, right caudate, left hippocampus, right hippocampus, left pallidum, right pallidum, left putamen, right putamen, left thalamus, right thalamus, left ventricle, right ventricle, all structures}. In general, the registration result is pretty good.



(A) (B)

Figure 4.7: DICE score of each step or iteration in registration. (A) DICE score on the 5th subject. (B) Average DICE score on the whole dataset.

Two examples of registration results are demonstrated in figure 4.8 with one from healthy control group and the other one from disease group. The transformed template images and segmentation images are shown. The Jacobian and norm field of

the transformation are demonstrated as well.

4.2.2 Classification Result

Given the geometric and photometric features were extracted, classification can be carried out. Two widely used classification algorithms are used here, i.e., linear SVM and PCA followed by LDA (PCA+LDA).

Result on left hippocampus

We first show the results achieved on hippocampus. The motivation of this experiment is to show the photometric and geometric features are complementary. So before mixing up all structures, the geometric features, photometric features extracted from left hippocampus and their combination are compared in terms of classification accuracy. Table 4.2 shows the results using different classification and/or feature selection methods and different features. A 10-fold cross-validation was repeated 10 times with random splits to evaluate the performance. The column "pho" and "geo" refer to "photometric features" and "geometric features". Photometric features are the intensity value of every voxel. Geometric features are the determinant of Jacobian matrix of deformation at each voxel. "TTFS" and "RFFS" indicate two feature selection methods, i.e. T-test and random forest. The column "rigid pho" refers to photometric features extracted from images after rigid registration (without lddmm registration). The column "pho+geo" is the result achieved by combining photo-

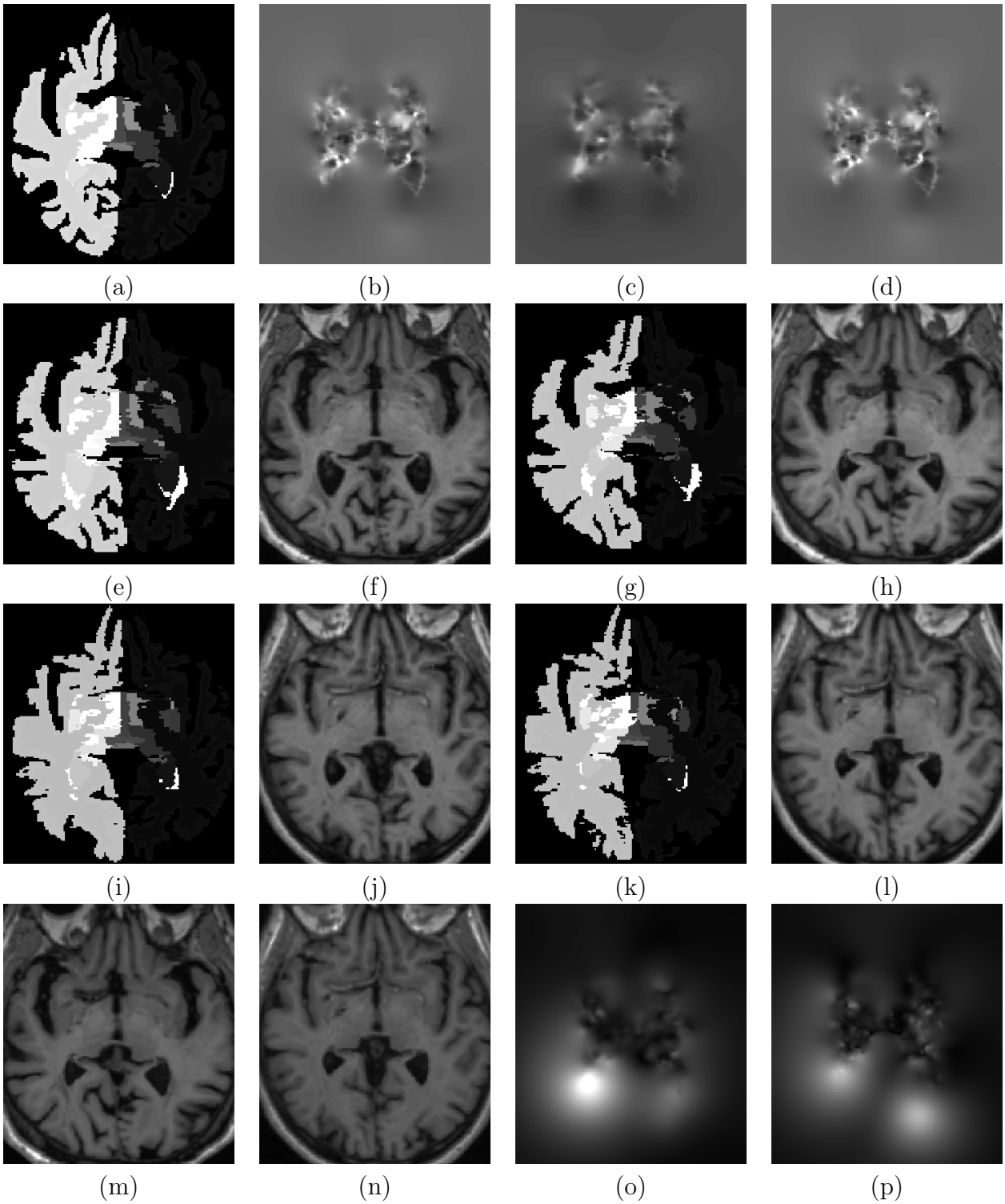


Figure 4.8: Example of registration result. (a) Segmentation image of the target. (b)∼(d) Three examples of Jacobian field. (e)∼(h) An example of healthy control subject, from left to right: segmentation and T1 image after affine registration, segmentation and T1 image after LDDMM registration. (i)∼(l) An example of diseased subject. (m)∼(n) Two examples of T1 images after histogram specification. (o)∼(p) Two examples of the field of norm of deformation.

metric and geometric features. Specifically we concatenated two feature vectors to combine them together.

One phenomena is that a higher accuracy can be achieved by the combination of geometric and photometric features. This implies these two features encode complementary information on one structure. Another thing we can observe is the feature selection does not benefit the classification in this case.

Table 4.2: Experimental result summary for left hippocampus. Here we keep the first K dimensions to retain 95% of the variance.

Method	"pho"	"pho rigid"	"geo"	"pho+geo"
Linear SVM	82.94%	76.12%	79.79%	84.68%
PCA+LDA	81.63%	75.59%	83.20%	84.78%
Linear SVM (TTFS)	81.51%	77.36%	80.15%	83.47%
PCA+LDA (TTFS)	82.63%	75.51%	81.20%	84.25%
Linear SVM (RFFS)	82.59%	78.64%	79.08%	84.04%
PCA+LDA (RFFS)	82.17%	75.72%	82.26%	84.51%

Results on All Structures

We extend the experiment from a single structure (the left hippocampus) to all structures (7 structures on both sides). The dimensionality increases dramatically to over 80,000 when all structures are taken into account. Classification algorithm like SVM can handle such kind of high dimension classification problem. For other methods like LDA it is necessary to do dimension reduction. We first show the visualization of the result of principal component analysis and T-test in figure 4.9. One can observe that the boundary regions have higher variance than inside region. This

is because the mismatching error occurs on the boundary. Two classification meth-

Figure 4.9: Visualization of the results of PCA and T-test. Left: a cross intersection of segmentation image. Middle: p-value of T-test. Right: the first principal component calculated using PCA

ods were attempted here. For all classification methods and features, we evaluated the performance using a 10-fold cross-validation, which was repeated 10 times with random splits. Feature selection before classification is an optional step. Here we tried random forest based feature selection (RFFS) and T-test based feature selection(TTFS). Firstly the results using LDA is presented in figure 4.10 and table 4.3. For almost all of structures, a higher performance can be observed for combination of geometric features and photometric features than using photometric features alone.

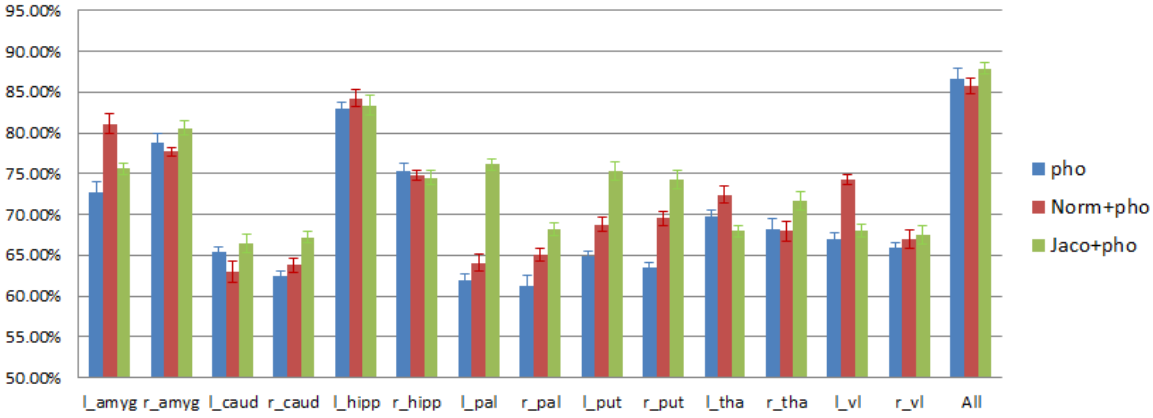


Figure 4.10: Accuracy achieved using PCA+LDA on different features.

Table 4.3: result of PCA+LDA (All Structures)

Features	Accuracy
Photometric features	86.60%±1.21%
Photometric features+norm of deformation	83.99%±0.99%
Photometric features+Jacobian of the mapping	86.93%±0.62%
Photometric features+Jacobian of the mapping(RFFS)	87.61%±0.79%
Photometric features+Jacobian of the mapping(TTFS)	86.10%±1.54%

Secondly the linear SVM was tested. As figure 4.11 and table 4.4 show, the approach of combining of geometric and photometric features outperforms using any of them individually. According to the results shown here, the classification performance can be improved by adding a step of random forest based feature selection. This phenomenon is not observed in experiments on the left hippocampus. This implies the feature selection methods tend to work better when dimensionality is large.

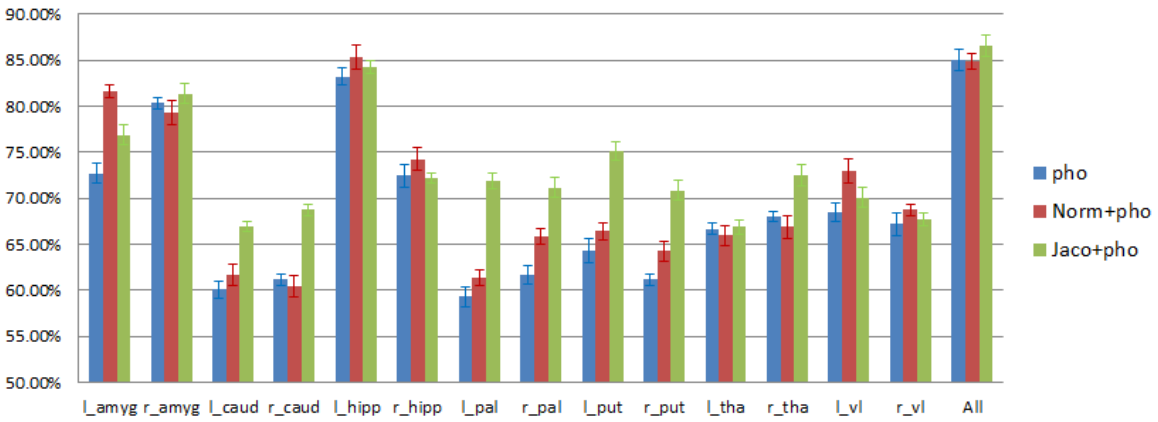


Figure 4.11: Accuracy achieved using linear SVM on different features.

Table 4.4: result of Linear SVM (All Structures)

Features	Accuracy
Photometric features	85.04% ±1.19%
Photometric features+norm of deformation	84.25% ±0.87%
Photometric features+Jacobian of the mapping	85.77%±0.41%
Photometric features+Jacobian of the mapping(RFFS)	86.53%±1.29%
Photometric features+Jacobian of the mapping(TTFS)	85.02%±2.04%

4.2.3 Results of Data Fusion Experiments

The results of three data fusion methods, majority voting, two-level SVM, and MK-SVM, are presented in table 4.5. Note that for all of these three methods, the random forest based feature selection was carried out to reduce the dimensionality.

Table 4.5: Results of data fusion methods

Method	Accuracy
Majority Voting	82.41% ±0.83%
Two-level SVM	86.17% ±1.85%
MK-SVM	88.25% ±1.17%

4.2.4 Comparison to state-of-the-art methods

As shown above, the best result is accomplished using MK-SVM. All experiments described so far are based on the subset of ADNI database with 210 healthy controls and 175 AD patients. We demonstrate the efficacy of our method by comparing the result of MK-SVM with two state-of-the-art approaches [18, 25]. But both of the two methods were validated on a smaller subset of ADNI database with 51 AD patients

and 52 healthy controls. Detailed subject information and the list of subject ID are provided in [18]. We compare these two methods to our method using the Jacobian field as geometric features and MK-SVM for data fusion and classification.

From table 4.6 we can see our method can achieve better or comparable performance compared with state-of-the-art approaches. The results listed here are those achieved only using MR images. Besides the two methods we compared with our method are both analysis of the whole brain MR scans. But our method only uses several regions of MR images of the human brain.

Table 4.6: Comparison our method with two state-of-the-art methods

Method	Accuracy	Sensitivity	Specificity
Our Method	88.4%±1.03%	88.2%±1.61%	88.5%±1.42%
[18]	86.20%	86.0%	86.3%
[25]	89.1%	89.2%	89.0%

4.3 Conclusion and Discussion

In this chapter, we presented a framework incorporating the feature extraction, fusion and inference of geometric and functional contrast of MR images. The MC-LDDMM was utilized for non-linear alignment. The determinant of Jacobian matrix at each position was used as geometric feature. The intensity value of each voxel presents the functional contrast of the brain. A binary Classification between AD and HC groups was carried out on these two features. A variety of feature extraction

and selection techniques were employed to reduce the dimensionality. According to the results reported above this framework can achieve comparable or even better classification accuracy rate when compared with two state-of-the-art methods.

The registration result affects the final classification performance. Thus a multiple-round MC-LDDMM was employed. This registration is the most time consuming step in our framework. The typical running time of our MC-LDDMM is approximately 5 hours. The timing is based on one cluster node of a IBM iDataPlex cluster with 8 cores provided by the Institute for Computational Medicine (ICM). This running time is acceptable for a off-line analysis.

In the future we will explore how to extract and select various features taking spatial information into account. For example some local features like Haar like feature [73] or local binary pattern (LBP) [74] could be explored. Such kind of patch based analysis has been proven to be effective for AD detection [75, 76].

Chapter 5

Conclusion and Future Work

5.1 Conclusion

This dissertation has discussed the fusion and inference of geometric information and functional contrast in computational anatomy framework. The application of the idea for Alzheimer's disease diagnosis has been demonstrated.

In chapter 1, a manifold learning based method was presented and compared with similar methods. According to the result achieved on the same database, our method could achieve a better or comparable performance in terms of prediction accuracy. But different from related work, we validated our method on large scale dataset. Furthermore, the analysis was only based on a single structure instead of the whole brain. However a comparable result was achieved on a single structure compared with related works.

Second, multiple structure analysis was presented in chapter 2. A dissimilarity measurement was calculated and transformed into a kernel matrix. Multi-kernel SVM was utilized to find the optimal linear combination of kernel matrices generated from different structures. The result shows this is an effective way for data fusion of different subcortical structures. A higher accuracy can be achieved using Multiple structure analysis than the single structure analysis in chapter 1.

In chapter 3, intensity value of MR images are used for discriminant analysis as well. A framework was created to extract geometric features and functional contrast features, i.e. the intensity value of MR images. A variety of feature extraction and selection techniques were employed to reduce the dimensionality. Experimental results show this framework significantly outperforms the two methods only using shape information as described in chapter 1 and 2. When compared with two state-of-art methods this framework can achieve comparable or even better classification accuracy rate .

In sum, three ROI based analysis methods were described and validated on the ADNI database. Experimental results demonstrate that these methods can effectively extract various of features from human brain MR images and make inference based on these features. These methods are not limited to the diagnosis of Alzheimer's disease, but could be applied to detect all other anatomical abnormalities of human brain.

5.2 Future Work

For the surface based analysis described in chapter 2 and 3. A potential work is to involve more data. Especially those data without label (diagnosis). Along with those well labeled images from public database, the performance of the learning methods could be benefited. This kind of data with no label or unreliable label can be used directly in unsupervised learning and semi-supervised learning. The lack of data, especially labeled data is a very challenging problem in medical image analysis field. The reason is that medical images cannot be captured and collected as easy as nature images.

Another worthwhile future work is to generalize the knowledge we learned from one database to a new database. This is very important because currently most of the applications of medical image analysis is specified to a certain dataset. This problem is very challenging because different datasets use different configurations for image acquisition. Such kind of variation is dominant over the morphological variation across the disease group and healthy control group.

All the experiments demonstrated in this dissertation were carried out on one database (ADNI). But it would be more valuable if one can train classifiers on one database and test them on another database. This is fairly meaningful given the fact that the scalability of medical image database is not high as usual.

Other potential future works include to extend to multi-modality analysis and patch-based analysis as described in the end of chapter 4.

Bibliography

[1] A. Qiu, M. I. Miller *et al.*, "Multi-structure network shape analysis via normal surface momentum maps," *NeuroImage*, vol. 42, no. 4, p. 1430, 2008.

[2] U. Grenander and M. Miller, "Computational anatomy: An emerging discipline," *Quarterly of applied mathematics*, vol. 56, no. 4, pp. 617–694, 1998.

[3] S. Mueller, M. Weiner, L. Thal, R. Petersen, C. Jack, W. Jagust, J. Trojanowski, A. Toga, and L. Beckett, "The alzheimers disease neuroimaging initiative," *Neuroimaging Clinics of North America*, vol. 15, no. 4, p. 869, 2005.

[4] B. Fischl, D. Salat, E. Busa, M. Albert, M. Dieterich, C. Haselgrove, A. van der Kouwe, R. Killiany, D. Kennedy, S. Klaveness *et al.*, "Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain," *Neuron*, vol. 33, no. 3, pp. 341–355, 2002.

[5] A. R. Khan, L. Wang, and M. F. Beg, "Freesurfer-initiated fully-automated subcortical brain segmentation in mri using large deformation diffeomorphic metric mapping," *NeuroImage*, vol. 41, no. 3, p. 735, 2008.

BIBLIOGRAPHY

[6] D. Holland, J. B. Brewer, D. J. Hagler, C. Fennema-Notestine, A. M. Dale, M. Weiner, L. Thal, R. Petersen, C. R. Jack, W. Jagust *et al.*, "Subregional neuroanatomical change as a biomarker for alzheimer's disease," *Proceedings of the National Academy of Sciences*, vol. 106, no. 49, pp. 20 954–20 959, 2009.

[7] D. Holland, L. K. McEvoy, R. S. Desikan, A. M. Dale, A. D. N. Initiative *et al.*, "Enrichment and stratification for predementia alzheimer disease clinical trials," *PloS one*, vol. 7, no. 10, p. e47739, 2012.

[8] D. Holland, L. K. McEvoy, and A. M. Dale, "Unbiased comparison of sample size estimates from longitudinal structural measures in adni," *Human brain mapping*, vol. 33, no. 11, pp. 2586–2602, 2012.

[9] W. E. Lorensen and H. E. Cline, "Marching cubes: A high resolution 3d surface construction algorithm," in *ACM Siggraph Computer Graphics*, vol. 21, no. 4. ACM, 1987, pp. 163–169.

[10] A. Qiu, T. Brown, B. Fischl, J. Ma, and M. I. Miller, "Atlas generation for subcortical and ventricular structures with its applications in shape analysis," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1539–1547, 2010.

[11] M. Miller, A. Trouvé, and L. Younes, "Geodesic shooting for computational anatomy," *Journal of mathematical imaging and vision*, vol. 24, no. 2, pp. 209–228, 2006.

[12] X. Tang, D. Holland, A. M. Dale, L. Younes, and M. I. Miller, "Shape abnormalities of subcortical and ventricular structures in mild cognitive impairment and alzheimer's disease: detecting, quantifying, and predicting," *Human brain mapping*, 2014.

[13] M. I. Miller and L. Younes, "Group actions, homeomorphisms, and matching: A general framework," *International Journal of Computer Vision*, vol. 41, no. 1-2, pp. 61–84, 2001.

[14] P. Dupuis, U. Grenander, and M. I. Miller, "Variational problems on flows of diffeomorphisms for image matching," *Quarterly of applied mathematics*, vol. 56, no. 3, p. 587, 1998.

[15] A. Buja, D. F. Swayne, M. L. Littman, N. Dean, H. Hofmann, and L. Chen, "Data visualization with multidimensional scaling," *Journal of Computational and Graphical Statistics*, vol. 17, no. 2, pp. 444–472, 2008.

[16] R. W. Floyd, "Algorithm 97: shortest path," *Communications of the ACM*, vol. 5, no. 6, p. 345, 1962.

[17] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.

[18] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of

alzheimer's disease and mild cognitive impairment," *Neuroimage*, vol. 55, no. 3, pp. 856–867, 2011.

[19] R. Cuingnet, E. Gerardin, J. Tessieras, G. Auzias, S. Lehéricy, M.-O. Habert, M. Chupin, H. Benali, and O. Colliot, "Automatic classification of patients with alzheimer's disease from structural mri: a comparison of ten methods using the adni database," *Neuroimage*, vol. 56, no. 2, pp. 766–781, 2011.

[20] T. Tapiola, C. Pennanen, M. Tapiola, S. Tervo, M. Kivipelto, T. Hänninen, M. Pihlajamäki, M. P. Laakso, M. Hallikainen, A. Hämäläinen *et al.*, "Mri of hippocampus and entorhinal cortex in mild cognitive impairment: a follow-up study," *Neurobiology of aging*, vol. 29, no. 1, pp. 31–38, 2008.

[21] K. K. Leung, J. Barnes, G. R. Ridgway, J. W. Bartlett, M. J. Clarkson, K. Macdonald, N. Schuff, N. C. Fox, and S. Ourselin, "Automated cross-sectional and longitudinal hippocampal volume measurement in mild cognitive impairment and alzheimer's disease," *Neuroimage*, vol. 51, no. 4, pp. 1345–1359, 2010.

[22] E. Gerardin, G. Chételat, M. Chupin, R. Cuingnet, B. Desgranges, H.-S. Kim, M. Niethammer, B. Dubois, S. Lehéricy, L. Garnero *et al.*, "Multidimensional classification of hippocampal shape features discriminates alzheimer's disease and mild cognitive impairment from normal aging," *Neuroimage*, vol. 47, no. 4, pp. 1476–1486, 2009.

[23] O. Querbes, F. Aubry, J. Pariente, J.-A. Lotterie, J.-F. Démonet, V. Duret,

M. Puel, I. Berry, J.-C. Fort, P. Celsis *et al.*, "Early diagnosis of alzheimer's disease using cortical thickness: impact of cognitive reserve," *Brain*, vol. 132, no. 8, pp. 2036–2047, 2009.

[24] M. Lehmann, S. J. Crutch, G. R. Ridgway, B. H. Ridha, J. Barnes, E. K. Warrington, M. N. Rossor, and N. C. Fox, "Cortical thickness and voxel-based morphometry in posterior cortical atrophy and typical alzheimer's disease," *Neurobiology of aging*, vol. 32, no. 8, pp. 1466–1476, 2011.

[25] P. Zhang, C.-Y. Wee, M. Niethammer, D. Shen, and P.-T. Yap, "Large deformation image classification using generalized locality-constrained linear coding," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013*. Springer, 2013, pp. 292–299.

[26] P. Zhang, M. Niethammer, D. Shen, and P.-T. Yap, "Large deformation diffeomorphic registration of diffusion-weighted imaging data," *Medical Image Analysis*, 2014.

[27] M. Beg, M. Miller, A. Trouvé, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *International Journal of Computer Vision*, vol. 61, no. 2, pp. 139–157, 2005.

[28] H. Park, "Isomap induced manifold embedding and its application to alzheimer's disease and mild cognitive impairment," *Neuroscience Letters*, 2012.

[29] K. Gray, P. Aljabar, R. Heckemann, A. Hammers, and D. Rueckert, "Random forest-based manifold learning for classification of imaging data in dementia," *Machine Learning in Medical Imaging*, pp. 159–166, 2011.

[30] ——, "Random forest-based similarity measures for multi-modal classification of alzheimer's disease," *NeuroImage*, 2012.

[31] R. Wolz, P. Aljabar, J. Hajnal, J. Lötjönen, and D. Rueckert, "Nonlinear dimensionality reduction combining mr imaging with non-imaging information," *Medical Image Analysis*, 2011.

[32] X. Yang, A. Goh, and A. Qiu, "Approximations of the diffeomorphic metric and their applications in shape learning," in *Information Processing in Medical Imaging*. Springer, 2011, pp. 257–270.

[33] S. Gerber, T. Tasdizen, P. Fletcher, S. Joshi, and R. Whitaker, "Manifold modeling for brain population analysis," *Medical image analysis*, vol. 14, no. 5, p. 643, 2010.

[34] M. Miller, C. Priebe, A. Qiu, B. Fischl, A. Kolasny, T. Brown, Y. Park, J. Ratnanather, E. Busa, J. Jovicich *et al.*, "Collaborative computational anatomy: an mri morphometry study of the human brain via diffeomorphic metric mapping," *Human brain mapping*, vol. 30, no. 7, pp. 2132–2141, 2008.

[35] A. Rangarajan, H. Chui, and F. L. Bookstein, "The softassign procrustes match-

ing algorithm," in *Information Processing in Medical Imaging*. Springer, 1997, pp. 29–42.

[36] M. Vaillant, A. Qiu, J. Glaunès, and M. Miller, "Diffeomorphic metric surface mapping in subregion of the superior temporal gyrus," *NeuroImage*, vol. 34, no. 3, pp. 1149–1159, 2007.

[37] M. Vaillant and J. Glaunes, "Surface matching via currents," in *Information Processing in Medical Imaging*. Springer, 2005, pp. 1–5.

[38] M. W. Trosset, C. E. Priebe, Y. Park, and M. I. Miller, "Semisupervised learning from dissimilarity data," *Computational statistics & data analysis*, vol. 52, no. 10, pp. 4643–4657, 2008.

[39] J. Tenenbaum, V. De Silva, and J. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[40] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," *Advances in neural information processing systems*, vol. 14, pp. 585–591, 2001.

[41] A. Liaw and M. Wiener, "Classification and regression by randomforest," *R news*, vol. 2, no. 3, pp. 18–22, 2002.

[42] B. E. Boser, I. M. Guyon, and V. N. Vapnik, "A training algorithm for optimal

margin classifiers," in *Proceedings of the fifth annual workshop on Computational learning theory*. ACM, 1992, pp. 144–152.

[43] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, pp. 27:1–27:27, 2011, software available at http://www.csie.ntu.edu.tw/~cjlin/libsvm.

[44] ——, "Libsvm: a library for support vector machines," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 2, no. 3, p. 27, 2011.

[45] G. C. Cawley and N. L. Talbot, "On over-fitting in model selection and subsequent selection bias in performance evaluation," *The Journal of Machine Learning Research*, vol. 11, pp. 2079–2107, 2010.

[46] C. M. Bishop *et al.*, *Pattern recognition and machine learning*. springer New York, 2006, vol. 1.

[47] M. Vaillant, M. I. Miller, L. Younes, A. Trouvé *et al.*, "Statistics on diffeomorphisms via tangent space representations," *NeuroImage*, vol. 23, no. 1, p. 161, 2004.

[48] L. Wang, F. Beg, T. Ratnanather, C. Ceritoglu, L. Younes, J. C. Morris, J. G. Csernansky, and M. I. Miller, "Large deformation diffeomorphism and momentum based hippocampal shape discrimination in dementia of the alzheimer type," *Medical Imaging, IEEE Transactions on*, vol. 26, no. 4, pp. 462–470, 2007.

[49] T. Tong, R. Wolz, Q. Gao, R. Guerrero, J. V. Hajnal, and D. Rueckert, "Multiple instance learning for classification of dementia in brain mri," *Medical image analysis*, vol. 18, no. 5, pp. 808–818, 2014.

[50] S. Klein, M. Loog, F. van der Lijn, T. den Heijer, A. Hammers, M. de Bruijne, A. van der Lugt, R. P. Duin, M. M. Breteler, and W. J. Niessen, "Early diagnosis of dementia based on intersubject whole-brain dissimilarities," in *Biomedical Imaging: From Nano to Macro, 2010 IEEE International Symposium on*. IEEE, 2010, pp. 249–252.

[51] A. Luchtenberg, R. Simões, A.-M. v. C. van Walsum, and C. H. Slump, "Early detection of alzheimer's disease using histograms in a dissimilarity-based classification framework," in *SPIE Medical Imaging*. International Society for Optics and Photonics, 2014, pp. 903 502–903 502.

[52] C. E. Priebe, Z. Ma, D. J. Marchette, E. Hohman, and G. Coppersmith, "Fusion and inference from multiple data sources," *The 57th Session of the International Statistical Institute*, 2009.

[53] M. Sun and C. E. Priebe, "Efficiency investigation of manifold matching for text document classification," *Pattern Recognition Letters*, 2013.

[54] F. R. Bach, G. R. Lanckriet, and M. I. Jordan, "Multiple kernel learning, conic duality, and the smo algorithm," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 6.

[55] M. Gönen and E. Alpaydın, "Multiple kernel learning algorithms," *The Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, 2011.

[56] S. Vishwanathan, Z. sun, N. Ampornpunt, and M. Varma, "Multiple kernel learning and the smo algorithm." in *NIPS*, vol. 22, 2010, p. 2.

[57] D. Zhang and D. Shen, "Multi-modal multi-task learning for joint prediction of multiple regression and classification variables in alzheimer's disease," *Neuroimage*, vol. 59, no. 2, pp. 895–907, 2012.

[58] J. Ashburner, C. Hutton, R. Frackowiak, I. Johnsrude, C. Price, K. Friston *et al.*, "Identifying global anatomical differences: deformation-based morphometry," *Human brain mapping*, vol. 6, no. 5-6, pp. 348–357, 1998.

[59] J. Ashburner and K. J. Friston, "Voxel-based morphometrythe methods," *Neuroimage*, vol. 11, no. 6, pp. 805–821, 2000.

[60] R. P. Woods, S. T. Grafton, C. J. Holmes, S. R. Cherry, and J. C. Mazziotta, "Automated image registration: I. general methods and intrasubject, intramodality validation," *Journal of computer assisted tomography*, vol. 22, no. 1, pp. 139–152, 1998.

[61] R. P. Woods, S. T. Grafton, J. D. Watson, N. L. Sicotte, and J. C. Mazziotta, "Automated image registration: Ii. intersubject validation of linear and nonlinear

models," *Journal of computer assisted tomography*, vol. 22, no. 1, pp. 153–165, 1998.

[62] A. Klein, J. Andersson, B. A. Ardekani, J. Ashburner, B. Avants, M.-C. Chiang, G. E. Christensen, D. L. Collins, J. Gee, P. Hellier *et al.*, "Evaluation of 14 nonlinear deformation algorithms applied to human brain mri registration," *Neuroimage*, vol. 46, no. 3, pp. 786–802, 2009.

[63] C. Ceritoglu, K. Oishi, X. Li, M.-C. Chou, L. Younes, M. Albert, C. Lyketsos, P. C. van Zijl, M. I. Miller, and S. Mori, "Multi-contrast large deformation diffeomorphic metric mapping for diffusion tensor imaging," *Neuroimage*, vol. 47, no. 2, p. 618, 2009.

[64] M.-C. Chiang, R. A. Dutton, K. M. Hayashi, O. L. Lopez, H. J. Aizenstein, A. W. Toga, J. T. Becker, and P. M. Thompson, "3d pattern of brain atrophy in hiv/aids visualized using tensor-based morphometry," *Neuroimage*, vol. 34, no. 1, pp. 44–60, 2007.

[65] A. Qiu, C. Fennema-Notestine, A. M. Dale, and M. I. Miller, "Regional shape abnormalities in mild cognitive impairment and alzheimer's disease," *Neuroimage*, vol. 45, no. 3, pp. 656–661, 2009.

[66] Y. Wang, Y. Song, P. Rajagopalan, T. An, K. Liu, Y.-Y. Chou, B. Gutman, A. W. Toga, and P. M. Thompson, "Surface-based tbm boosts power to detect

disease effects on the brain: an n= 804 adni study," *Neuroimage*, vol. 56, no. 4, pp. 1993–2010, 2011.

[67] R. P. Woods, "Characterizing volume and surface deformations in an atlas framework: theory, applications, and implementation," *NeuroImage*, vol. 18, no. 3, pp. 769–788, 2003.

[68] R. Genuer, J.-M. Poggi, and C. Tuleau-Malot, "Variable selection using random forests," *Pattern Recognition Letters*, vol. 31, no. 14, pp. 2225–2236, 2010.

[69] M. Sandri and P. Zuccolotto, "Variable selection using random forests," in *Data analysis, classification and the forward search.* Springer, 2006, pp. 263–270.

[70] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis.* John Wiley & Sons, 2004, vol. 46.

[71] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," *Geoscience and Remote Sensing, IEEE Transactions on*, vol. 45, no. 12, pp. 3858–3866, 2007.

[72] S. Yu, L.-C. Tranchevent, B. Moor, and Y. Moreau, *Kernel-based data fusion for machine learning: methods and applications in bioinformatics and text mining.* Springer, 2011, vol. 345.

[73] E. Bicacro, M. Silveira, and J. S. Marques, "Alternative feature extraction methods in 3d brain image-based diagnosis of alzheimer's disease," in *Image Process-*

ing (ICIP), 2012 19th IEEE International Conference on. IEEE, 2012, pp. 1237–1240.

[74] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, no. 7, pp. 971–987, 2002.

[75] B. Babenko, "Multiple instance learning: algorithms and applications," 2008.

[76] T. Tong, R. Wolz, Q. Gao, J. V. Hajnal, and D. Rueckert, "Multiple instance learning for classification of dementia in brain mri," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013.* Springer, 2013, pp. 599–606.

Vita

Jianqiao Feng received the En. B. degree in Computer Science from Shandong University in 2006, and Sc. Master degree in Computer Science from Peking University in 2009. He enrolled in Ph.D. program at Johns Hopkins University in 2009. His research focuses machine learning application on medical image analysis, and computer aided diagnosis.